

# Science Collection

## Alternative Science

# Contents

## Articles

Science	1
Natural science	21
Scientific method	32
DNA	59
Computer	83
Antenna (radio)	102
Brain	123
Neuron	146
Fractal	160
Fibonacci number	172
Golden ratio	188
Introduction to quantum mechanics	210
Quantum Zeno effect	229
Observer effect (physics)	234
Uncertainty principle	235
Quantum tunnelling	257
Implicate and explicate order according to David Bohm	266
Biocentrism (theory of everything)	272
Fine-tuned Universe	275
Conservation of energy	283
Zero-point energy	289
Vacuum energy	294
Casimir effect	297
Consciousness	306
Stuart Hameroff	325
Roger Penrose	328
Holography	337
Mathematics	357
Language	371
Collective unconscious	400
Jungian archetypes	402
Schumann resonances	409
Plasma (physics)	418
Space	434



Chronon	441
---------	-----

## References

Article Sources and Contributors	442
----------------------------------	-----

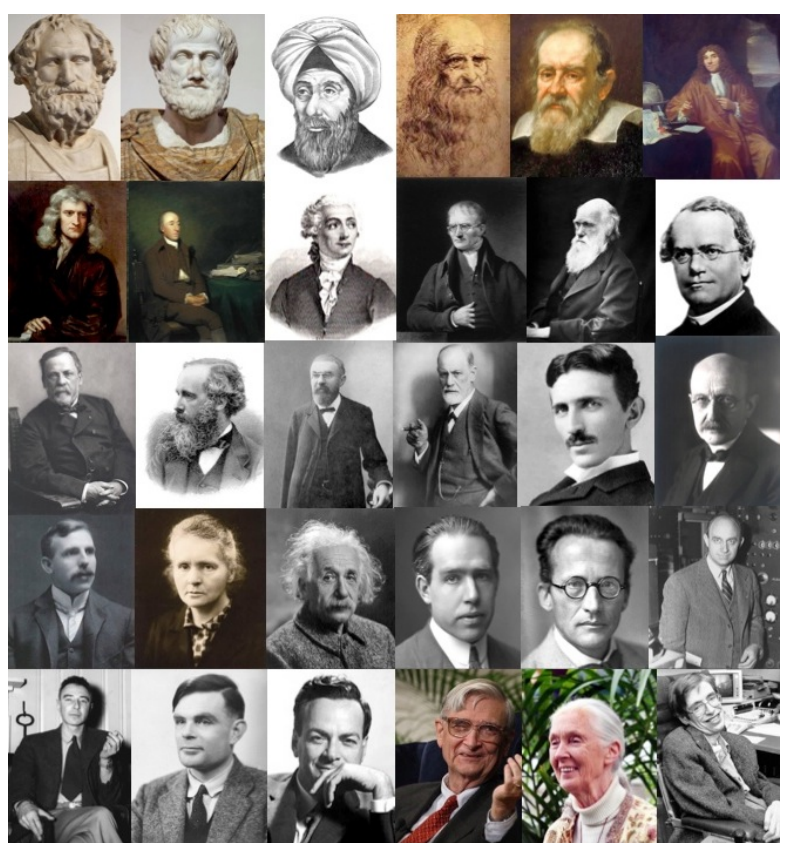
Image Sources, Licenses and Contributors	455
--	-----

## Article Licenses

License	463
---------	-----

# Science

**Science** (from Latin *scientia*, meaning "knowledge") is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe.<sup>[1]</sup> In an older and closely related meaning (found, for example, in Aristotle), "science" refers to the body of reliable knowledge itself, of the type that can be logically and rationally explained (*see History and philosophy below*).<sup>[2]</sup> Since classical antiquity science as a type of knowledge was closely linked to philosophy. In the early modern era the words "science" and "philosophy" were sometimes used interchangeably in the English language. By the 17th century, natural philosophy (which is today called "natural science") was considered a separate branch of philosophy.<sup>[3]</sup> However, "science" continued to be used in a broad sense denoting reliable knowledge about a topic, in the same way it is still used in modern terms such as library science or political science.



Montage of some highly influential scientists from a variety of scientific fields. From left to right: Top row: Archimedes, Aristotle, Ibn al-Haytham, Leonardo da Vinci, Galileo Galilei, Antonie van Leeuwenhoek; Second row: Isaac Newton, James Hutton, Antoine Lavoisier, John Dalton, Charles Darwin, Gregor Mendel; Third row: Louis Pasteur, James Clerk Maxwell, Henri Poincaré, Sigmund Freud, Nikola Tesla, Max Planck; Fourth row: Ernest Rutherford, Marie Curie, Albert Einstein, Niels Bohr, Erwin Schrödinger, Enrico Fermi; Bottom row: J. Robert Oppenheimer, Alan Turing, Richard Feynman, E. O. Wilson, Jane Goodall, Stephen Hawking

In modern use, "science" more often refers to a way of pursuing knowledge, not only the knowledge itself. It is "often treated as synonymous with 'natural and physical science', and thus restricted to those branches of study that relate to the phenomena of the material universe and their laws, sometimes with implied exclusion of pure mathematics. This is now the dominant sense in ordinary use."<sup>[4]</sup> This narrower sense of "science" developed as scientists such as Johannes Kepler, Galileo Galilei and Isaac Newton began formulating *laws of nature* such as Newton's laws of motion. In this period it became more common to refer to natural philosophy as "natural science". Over the course of the 19th century, the word "science" became increasingly associated with the scientific method, a disciplined way to study the natural world, including physics, chemistry, geology and biology. It is in the 19th century also that the term *scientist* was created by the naturalist-theologian William Whewell to distinguish those who sought knowledge on nature from those who sought knowledge on other disciplines. The *Oxford English Dictionary* dates the origin of the word "scientist" to 1834. This sometimes left the study of human thought and society in a linguistic limbo, which was resolved by classifying these areas of academic study as social science. Similarly, several other major areas of disciplined study and knowledge exist today under the general rubric of "science", such as formal science and applied science.

## History and philosophy

## History

Science in a broad sense existed before the modern era, and in many historical civilizations, but modern science is so distinct in its approach and successful in its results that it now defines what science is in the strictest sense of the term. Much earlier than the modern era, another important turning point was the development of the classical natural philosophy in the ancient Greek-speaking world.

### Pre-philosophical

Science in its original sense is a word for a type of knowledge (Latin *scientia*, Ancient Greek *epistemē*), rather than a specialized word for the pursuit of such knowledge. In particular it is one of the types of knowledge which people can communicate to each other and share. For example, knowledge about the working of natural things was gathered long before recorded history and led to the development of complex abstract thinking, as shown by the construction of complex calendars, techniques for making poisonous plants edible, and buildings such as the pyramids. However no consistent distinction was made between knowledge of such things which are true in every community, and other types of communal knowledge such as mythologies and legal systems.

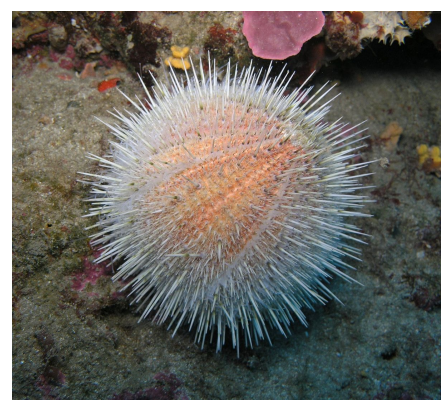
### Philosophical study of nature

Before the invention or discovery of the concept of "nature" (Ancient Greek *phusis*), by the Pre-Socratic philosophers, the same words tend to be used to describe the *natural* "way" in which a plant grows,<sup>[6]</sup> and the "way" in which, for example, one tribe worships a particular god. For this reason it is claimed these men were the first philosophers in the strict sense, and also the first people to clearly distinguish "nature" and "convention".<sup>[7]</sup> Science was therefore distinguished as the knowledge of nature, and the things which are true for every community, and the name of the specialized pursuit of such knowledge was philosophy — the realm of the first philosopher-physicists. They were mainly speculators or theorists, particularly interested in astronomy. In contrast, trying to use knowledge of nature to imitate nature (artifice or technology, Greek *technē*) was seen by classical scientists as a more appropriate interest for lower class artisans.<sup>[8]</sup>

### Philosophical turn to human things

A major turning point in the history of early philosophical science was the controversial but successful attempt by Socrates to apply philosophy to the study of human things, including human nature, the nature of political communities, and human knowledge itself. He criticized the older type of study of physics as too purely speculative, and lacking in self-criticism. He was particularly concerned that some of the early physicists treated nature as if it could be assumed that it had no intelligent order, explaining things merely in terms of motion and matter.

The study of human things had been the realm of mythology and tradition, and Socrates was executed. Aristotle later created a less controversial systematic programme of Socratic philosophy, which was teleological, and human-centred. He rejected many of the conclusions of earlier scientists. For example in his physics the sun goes around the earth, and many things have it as part of their nature that they are for humans. Each thing has a formal cause and final cause and a role in the rational cosmic order. Motion and change is described as the actualization of potentials already in things, according to what types of things they are. While the Socratics insisted that philosophy should be used to consider the practical question of the best way to live for a human being, they did not argue for any



Both Aristotle and Kuan Tzu (4th C. BCE), in an example of simultaneous scientific discovery, mention that some marine animals were subject to a lunar cycle, and increase and decrease in size with the waxing and waning of the moon. Aristotle was referring specifically to the sea urchin, pictured above.<sup>[5]</sup>

other types of applied science.

Aristotle maintained the sharp distinction between science and the practical knowledge of artisans, treating theoretical speculation as the highest type of human activity, practical thinking about good living as something less lofty, and the knowledge of artisans as something only suitable for the lower classes. In contrast to modern science, Aristotle's influential emphasis was upon the "theoretical" steps of deducing universal rules from raw data, and did not treat the gathering of experience and raw data as part of science itself.<sup>[9]</sup>

### Medieval science

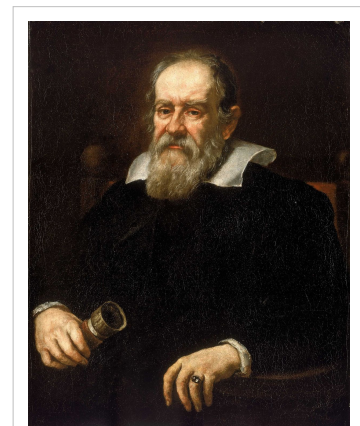
During late antiquity and the early Middle Ages, the Aristotelian approach to inquiries on natural phenomenon was used. Some ancient knowledge was lost, or in some cases kept in obscurity, during the fall of the Roman Empire and periodic political struggles. However, the general fields of science, or Natural Philosophy as it was called, and much of the general knowledge from the ancient world remained preserved through the works of the early encyclopedists like Isidore of Seville. During the early medieval period, Syrian Christians from Eastern Europe such as Nestorians and Monophysites were the ones that translated much of the important Greek science texts from Greek to Syriac and the later on they translated many of the works into Arabic and other languages under Islamic rule.<sup>[10]</sup> This was a major line of transmission for the development of Islamic science which provided much of the activity during the early medieval period. In the later medieval period, Europeans recovered some ancient knowledge by translations of texts and they built their work upon the knowledge of Aristotle, Ptolemy, Euclid, and others works. In Europe, men like Roger Bacon learned Arabic and Hebrew and argued for more experimental science. By the late Middle Ages, a synthesis of Catholicism and Aristotelianism known as Scholasticism was flourishing in Western Europe, which had become a new geographic center of science.

### Renaissance, and early modern science

By the late Middle Ages, especially in Italy there was an influx of texts and scholars from the collapsing Byzantine empire. Copernicus formulated a heliocentric model of the solar system unlike the geocentric model of Ptolemy's *Almagest*. All aspects of scholasticism were criticized in the 15th and 16th centuries; one author who was notoriously persecuted was Galileo, who made innovative use of experiment and mathematics. However the persecution began after Pope Urban VIII blessed Galileo to write about the Copernican system. Galileo had used arguments from the Pope and put them in the voice of the simpleton in the work "Dialogue Concerning the Two Chief World Systems" which caused great offense to him.<sup>[12]</sup>

In Northern Europe, the new technology of the printing press was widely used to publish many arguments including some that disagreed with church dogma. René Descartes and Francis Bacon published philosophical arguments in favor of a new type of non-Aristotelian science. Descartes argued that mathematics could be used in order to study nature, as Galileo had done, and Bacon emphasized the importance of experiment over contemplation. Bacon also argued that science should aim for the first time at practical inventions for the improvement of all human life.

Bacon questioned the Aristotelian concepts of formal cause and final cause, and promoted the idea that science should study the laws of "simple" natures, such as heat, rather than assuming that there is any specific nature, or "formal cause", of each complex type of thing. This new modern science began to see itself as describing "laws of nature". This updated approach to studies in nature was seen as mechanistic.



Galileo is considered one of the fathers of modern science.<sup>[11]</sup>

## Age of Enlightenment

In the 17th and 18th centuries, the project of modernity, as had been promoted by Bacon and Descartes, led to rapid scientific advance and the successful development of a new type of natural science, mathematical, methodically experimental, and deliberately innovative. Newton and Leibniz succeeded in developing a new physics, now referred to as Newtonian physics, which could be confirmed by experiment and explained in mathematics. Leibniz also incorporated terms from Aristotelian physics, but now being used in a new non-teleological way, for example "energy" and "potential". But in the style of Bacon, he assumed that different types of things all work according to the same general laws of nature, with no special formal or final causes for each type of thing.

It is, during this period that the word "science" gradually became more commonly used to refer to the pursuit of a type of knowledge, and especially knowledge of nature — coming close in meaning to the old term "natural philosophy".

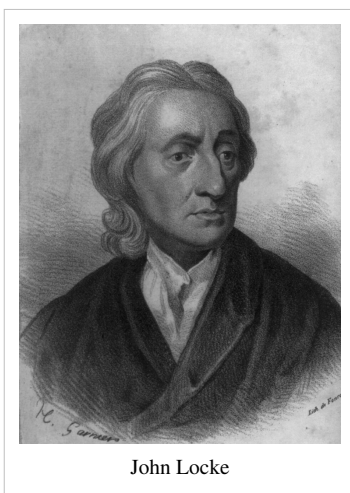
## 19th century

Both John Herschel and William Whewell systematised methodology: the latter coined the term scientist. When Charles Darwin published *On the Origin of Species* he established descent with modification as the prevailing evolutionary explanation of biological complexity. His theory of natural selection provided a natural explanation of how species originated, but this only gained wide acceptance a century later. John Dalton developed the idea of atoms. The laws of Thermodynamics and the electromagnetic theory were also established in the 19th century, which raised new questions which could not easily be answered using Newton's framework.

## 20th century and beyond

Einstein's Theory of Relativity and the development of quantum mechanics led to the replacement of Newtonian physics with a new physics which contains two parts, that describe different types of events in nature. The extensive use of scientific innovation during the wars of this century, led to the space race and widespread public appreciation of the importance of modern science. More recently it has been argued that the ultimate purpose of science is to make sense of human beings and our nature- for example in his book *Consilience*, EO Wilson said "The human condition is the most important frontier of the natural sciences."<sup>[13]</sup> Jeremy Griffith supports this view.<sup>[14]</sup>

## Philosophy of science



John Locke

Working scientists usually take for granted a set of basic assumptions that are needed to justify the scientific method: (1) that there is an objective reality shared by all rational observers; (2) that this objective reality is governed by natural laws; (3) that these laws can be discovered by means of systematic observation and experimentation. Philosophy of science seeks a deep understanding of what these underlying assumptions mean and whether they are valid.

The belief that all observers share a common reality is known as realism. It can be contrasted with anti-realism, the belief that there is no valid concept of absolute truth such that things that are true for one observer are true for all observers. The most commonly defended form of anti-realism is idealism, the belief that the mind or spirit is the most basic essence, and that each mind generates its own reality.<sup>[15]</sup> In an idealistic world-view, what is true for one

mind need not be true for other minds.

There are different schools of thought in philosophy of science. The most popular position is empiricism, which claims that knowledge is created by a process involving observation and that scientific theories are the result of generalizations from such observations.<sup>[16]</sup> Empiricism generally encompasses inductivism, a position that tries to



explain the way general theories can be justified by the finite number of observations humans can make and the hence finite amount of empirical evidence available to confirm scientific theories. This is necessary because the number of predictions those theories make is infinite, which means that they cannot be known from the finite amount of evidence using deductive logic only. Many versions of empiricism exist, with the predominant ones being bayesianism<sup>[17]</sup> and the hypothetico-deductive method.<sup>[18]</sup>

Empiricism has stood in contrast to rationalism, the position originally associated with Descartes, which holds that knowledge is created by the human intellect, not by observation.<sup>[21]</sup>

A significant twentieth century version of rationalism is critical rationalism, first defined by Austrian-British philosopher Karl Popper. Popper rejected the way that empiricism describes the connection between theory and observation. He claimed that theories are not generated by observation, but that observation is made in the light of theories and that the only way a theory can be affected by observation is when it comes in conflict with it.<sup>[22]</sup>

Popper proposed falsifiability as the landmark of scientific theories, and falsification as the empirical method, to replace verifiability<sup>[23]</sup> and induction by purely deductive notions.<sup>[24]</sup> Popper further claimed that there is actually only one universal method, and that this method is not specific to science: The negative method of criticism, trial and error.<sup>[25]</sup> It covers all products of the human mind, including science, mathematics, philosophy, and art<sup>[26]</sup>

Another approach, instrumentalism, colloquially termed "shut up and calculate", emphasizes the utility of theories as instruments for explaining and predicting phenomena.<sup>[27]</sup> It claims that scientific theories are black boxes with only their input (initial conditions) and output (predictions) being relevant. Consequences, notions and logical structure of the theories are claimed to be something that should simply be ignored and that scientists shouldn't make a fuss about (see interpretations of quantum mechanics).

Paul K Feyerabend advanced the idea of epistemological anarchism, which holds that there are no useful and exception-free methodological rules governing the progress of science or the growth of knowledge, and that the idea that science can or should operate according to universal and fixed rules is unrealistic, pernicious and detrimental to science itself.<sup>[28]</sup> Feyerabend advocates treating science as an ideology alongside others such as religion, magic and mythology, and considers the dominance of science in society authoritarian and unjustified. He also contended (along with Imre Lakatos) that the demarcation problem of distinguishing science from pseudoscience on objective grounds is not possible and thus fatal to the notion of science running according to fixed, universal rules.<sup>[28]</sup> Feyerabend also stated that science does not have evidence for its philosophical precepts, particularly the notion of Uniformity of Law and the Uniformity of Process across time and space.<sup>[29]</sup>

Finally, another approach often cited in debates of scientific skepticism against controversial movements like "scientific creationism", is methodological naturalism. Its main point is that a difference between natural and supernatural explanations should be made, and that science should be restricted methodologically to natural explanations.<sup>[30]</sup> That the restriction is merely methodological (rather than ontological) means that science should not consider supernatural explanations itself, but should not claim them to be wrong either. Instead, supernatural

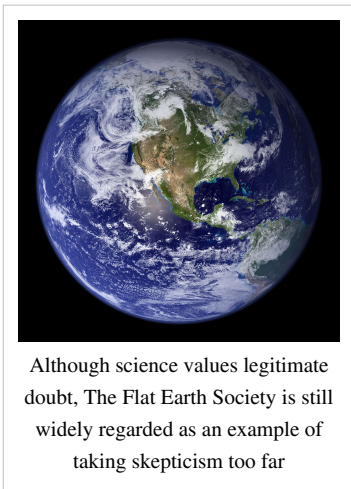
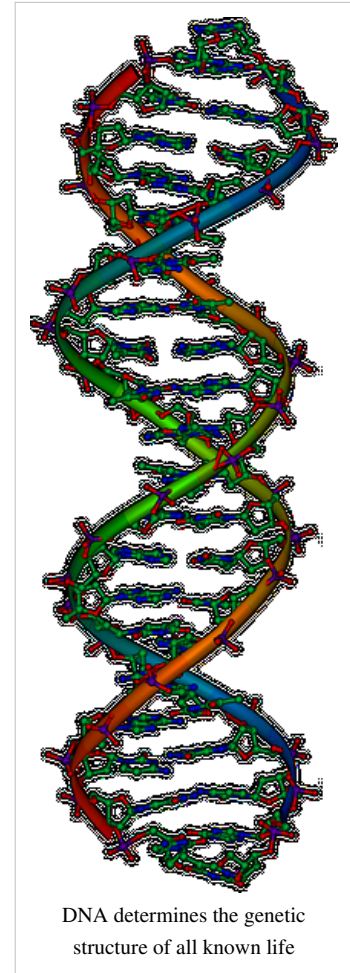


Distinguished Men of Science. Engraving after 'Men of Science Living in 1807-8'<sup>[19]</sup>, John Gilbert engraved by George Zobel and William Walker, ref. NPG 1075a, National Portrait Gallery, London, accessed February 2010 Use a cursor to see who is who. Smith, HM (May 1941). "Eminent men of science living in 1807-8". *J. Chem. Educ* 18 (5): 203. doi:10.1021/ed018p203. <http://pubs.acs.org/doi/abs/10.1021/ed018p203>.

explanations should be left a matter of personal belief outside the scope of science. Methodological naturalism maintains that proper science requires strict adherence to empirical study and independent verification as a process for properly developing and evaluating explanations for observable phenomena.<sup>[31]</sup> The absence of these standards, arguments from authority, biased observational studies and other common fallacies are frequently cited by supporters of methodological naturalism as criteria for the dubious claims they criticize not to be true science.

**Certainty and science**

A scientific theory is empirical, and is always open to falsification if new evidence is presented. That is, no theory is ever considered strictly certain as science accepts the concept of fallibilism. The philosopher of science Karl Popper sharply distinguishes truth from certainty. He writes that scientific knowledge "consists in the search for truth", but it "is not the search for certainty ... All human knowledge is fallible and therefore uncertain."<sup>[32]</sup>



New scientific knowledge very rarely results in vast changes in our understanding. According to psychologist Keith Stanovich, it may be the media's overuse of words like "breakthrough" that leads the public to imagine that science is constantly proving everything it thought was true to be false.<sup>[33]</sup> While there are such famous cases as the theory of relativity that required a complete reconceptualization, these are extreme exceptions. Knowledge in science is gained by a gradual synthesis of information from different experiments, by various researchers, across different domains of science; it is more like a climb than a leap.<sup>[34]</sup> Theories vary in the extent to which they have been tested and verified, as well as their acceptance in the scientific community.<sup>[35]</sup> For example, heliocentric theory, the theory of evolution, and germ theory still bear the name "theory" even though, in practice, they are considered factual.<sup>[36]</sup> Philosopher Barry Stroud adds that, although the best definition for "knowledge" is contested, being skeptical and entertaining the *possibility* that one is incorrect is compatible with being correct. Ironically then,

the scientist adhering to proper scientific approaches will doubt themselves even once they possess the truth.<sup>[37]</sup> The fallibilist C. S. Peirce argued that inquiry is the struggle to resolve actual doubt and that merely quarrelsome, verbal, or hyperbolic doubt is fruitless<sup>[38]</sup>—but also that the inquirer should try to attain genuine doubt rather than resting uncritically on common sense.<sup>[39]</sup> He held that the successful sciences trust, not to any single chain of inference (no stronger than its weakest link), but to the cable of multiple and various arguments intimately connected.<sup>[40]</sup>

Stanovich also asserts that science avoids searching for a "magic bullet"; it avoids the single-cause fallacy. This means a scientist would not ask merely "What is *the* cause of...", but rather "What *are* the most significant *causes* of...". This is especially the case in the more macroscopic fields of science (e.g. psychology, cosmology).<sup>[41]</sup> Of course, research often analyzes few factors at once, but these are always added to the long list of factors that are most important to consider.<sup>[41]</sup> For example: knowing the details of only a person's genetics, or their history and upbringing, or the current situation may not explain a behaviour, but a deep understanding of all these variables combined can be very predictive.

### **Pseudoscience, fringe science, and junk science**

An area of study or speculation that masquerades as science in an attempt to claim a legitimacy that it would not otherwise be able to achieve is sometimes referred to as pseudoscience, fringe science, or "alternative science".<sup>[42]</sup> Another term, junk science, is often used to describe scientific hypotheses or conclusions which, while perhaps legitimate in themselves, are believed to be used to support a position that is seen as not legitimately justified by the totality of evidence. Physicist Richard Feynman coined the term "cargo cult science" in reference to pursuits that have the formal trappings of science but lack "a principle of scientific thought that corresponds to a kind of utter honesty" that allows their results to be rigorously evaluated.<sup>[43]</sup> Various types of commercial advertising, ranging from hype to fraud, may fall into these categories.

There also can be an element of political or ideological bias on all sides of such debates. Sometimes, research may be characterized as "bad science", research that is well-intentioned but is seen as incorrect, obsolete, incomplete, or over-simplified expositions of scientific ideas. The term "scientific misconduct" refers to situations such as where researchers have intentionally misrepresented their published data or have purposely given credit for a discovery to the wrong person.<sup>[44]</sup>



## Scientific practice

"If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts, he shall end in certainties." —Francis Bacon (1605) *The Advancement of Learning*, Book 1, v, 8

A skeptical point of view, demanding a method of proof, was the practical position taken as early as 1000 years ago, with Alhazen, *Doubts Concerning Ptolemy*, through Bacon (1605), and C. S. Peirce (1839–1914), who note that a community will then spring up to address these points of uncertainty. The methods of inquiry into a problem have been known for thousands of years,<sup>[45]</sup> and extend beyond theory to practice. The use of measurements, for example, is a practical approach to settle disputes in the community.

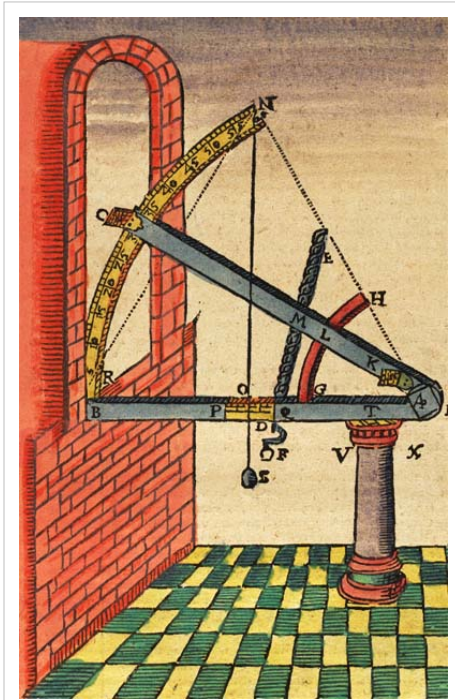
John Ziman points out that intersubjective pattern recognition is fundamental to the creation of all scientific knowledge.<sup>[46]</sup> Ziman shows how scientists can identify patterns to each other across centuries: Needham 1954 (illustration facing page 164) shows how today's trained Western botanist can identify *Artemisia alba* from images taken from a 16th c. Chinese pharmacopia,<sup>[47]</sup> and Ziman refers to this ability as 'perceptual consensibility'.<sup>[48]</sup> Ziman then makes consensibility, leading to consensus, the touchstone of reliable knowledge.<sup>[49]</sup>

## The scientific method

The scientific method seeks to explain the events of nature in a reproducible way.<sup>[50]</sup> An explanatory thought experiment or hypothesis is put forward, as explanation, using principles such as parsimony (also known as "Occam's Razor") and are generally expected to seek consilience—fitting well with other accepted facts related to the phenomena.<sup>[51]</sup> This new explanation is used to make falsifiable predictions that are testable by experiment or observation. The predictions are to be posted before a confirming experiment or observation is sought, as proof that no tampering has occurred. Disproof of a prediction is evidence of progress.<sup>[52][53]</sup> This is done partly through observation of natural phenomena, but also through experimentation, that tries to simulate natural events under controlled conditions, as appropriate to the discipline (in the observational sciences, such as astronomy or geology, a predicted observation might take the place of a controlled experiment). Experimentation is especially important in science to help establish causal relationships (to avoid the correlation fallacy).

When a hypothesis proves unsatisfactory, it is either modified or discarded.<sup>[54]</sup> If the hypothesis survived testing, it may become adopted into the framework of a scientific theory. This is a logically reasoned, self-consistent model or framework for describing the behavior of certain natural phenomena. A theory typically describes the behavior of much broader sets of phenomena than a hypothesis; commonly, a large number of hypotheses can be logically bound together by a single theory. Thus a theory is a hypothesis explaining various other hypotheses. In that vein, theories are formulated according to most of the same scientific principles as hypotheses. In addition to testing hypotheses, scientists may also generate a model based on observed phenomena. This is an attempt to describe or depict the phenomenon in terms of a logical, physical or mathematical representation and to generate new hypotheses that can be tested.<sup>[55]</sup>

While performing experiments to test hypotheses, scientists may have a preference for one outcome over another, and so it is important to ensure that science as a whole can eliminate this bias.<sup>[56][57]</sup> This can be achieved by careful



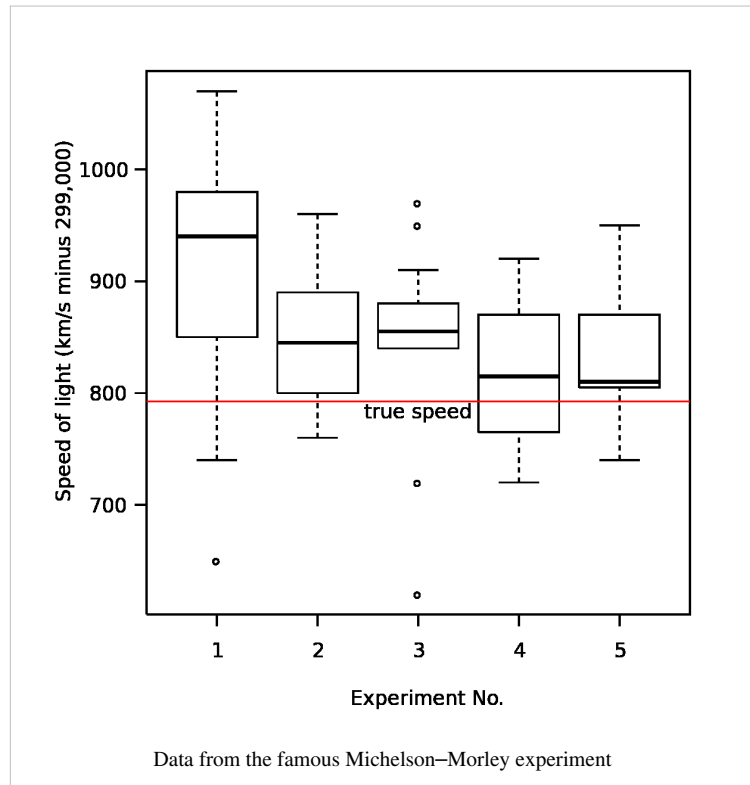
Astronomy became much more accurate after Tycho Brahe devised his scientific instruments for measuring angles between two celestial bodies, before the invention of the telescope. Brahe's observations were the basis for Kepler's laws.

experimental design, transparency, and a thorough peer review process of the experimental results as well as any conclusions.<sup>[58][59]</sup> After the results of an experiment are announced or published, it is normal practice for independent researchers to double-check how the research was performed, and to follow up by performing similar experiments to determine how dependable the results might be.<sup>[60]</sup> Taken in its entirety, the scientific method allows for highly creative problem solving while minimizing any effects of subjective bias on the part of its users (namely the confirmation bias).<sup>[61]</sup>

## Mathematics and formal sciences

Mathematics is essential to the sciences. One important function of mathematics in science is the role it plays in the expression of scientific models. Observing and collecting measurements, as well as hypothesizing and predicting, often require extensive use of mathematics. Arithmetic, algebra, geometry, trigonometry and calculus, for example, are all essential to physics. Virtually every branch of mathematics has applications in science, including "pure" areas such as number theory and topology.

Statistical methods, which are mathematical techniques for summarizing and analyzing data, allow scientists to assess the level of reliability and the range of variation in experimental results. Statistical analysis plays a fundamental role in many areas of both the natural sciences and social sciences.



Computational science applies computing power to simulate real-world situations, enabling a better understanding of scientific problems than formal mathematics alone can achieve. According to the Society for Industrial and Applied Mathematics, computation is now as important as theory and experiment in advancing scientific knowledge.<sup>[62]</sup>

Whether mathematics itself is properly classified as science has been a matter of some debate. Some thinkers see mathematicians as scientists, regarding physical experiments as inessential or mathematical proofs as equivalent to experiments. Others do not see mathematics as a science, since it does not require an experimental test of its theories and hypotheses. Mathematical theorems and formulas are obtained by logical derivations which presume axiomatic systems, rather than the combination of empirical observation and logical reasoning that has come to be known as the scientific method. In general, mathematics is classified as formal science, while natural and social sciences are classified as empirical sciences.<sup>[63]</sup>

## Basic and applied research

Although some scientific research is applied research into specific problems, a great deal of our understanding comes from the curiosity-driven undertaking of basic research. This leads to options for technological advance that were not planned or sometimes even imaginable. This point was made by Michael Faraday when, allegedly in response to the question "what is the *use* of basic research?" he responded "Sir, what is the use of a new-born child?".<sup>[64]</sup> For example, research into the effects of red light on the human eye's rod cells did not seem to have any practical purpose; eventually, the discovery that our night vision is not troubled by red light would lead search and rescue teams (among others) to adopt red light in the cockpits of jets and helicopters.<sup>[65]</sup> In a nutshell: Basic research is the search for knowledge. Applied research is the search for solutions to practical problems using this knowledge. Finally, even basic research can take unexpected turns, and there is some sense in which the scientific method is built to harness luck.

## Research in practice

Due to the increasing complexity of information and specialization of scientists, most of the cutting-edge research today is done by well funded groups of scientists, rather than individuals.<sup>[66]</sup> D.K. Simonton notes that due to the breadth of very precise and far reaching tools already used by researchers today and the amount of research generated so far, creation of new disciplines or revolutions within a discipline may no longer be possible as it is unlikely that some phenomenon that merits its own discipline has been overlooked. Hybridizing of disciplines and finessing knowledge is, in his view, the future of science.<sup>[66]</sup>

## Practical impacts of scientific research

Discoveries in fundamental science can be world-changing, but often take time to have that effect. For example:

Research	Impact
The strange orbit of Mercury and other research leading to special and general relativity	Satellite-based technology such as GPS, satnav and satellite communications <sup>[67]</sup>
Radioactivity and antimatter	Cancer treatment, PET scans, and medical research (via isotopic labeling)
Immunology	Vaccination, leading to the elimination of most infectious diseases from developed countries and the worldwide eradication of smallpox; hygiene, leading to decreased transmission of infectious diseases; antibodies, leading to techniques for disease diagnosis and targeted anticancer therapies.
Crystallography and quantum mechanics	Semiconductor devices, hence modern computing and telecommunications including the integration with wireless devices: the mobile phone <sup>[67]</sup>
Diffraction	Optics, hence fiberoptic cable, modern intercontinental communications, and cable TV and internet
Photovoltaic effect	Solar cells, hence solar power, solar powered watches, calculators and other devices.
Radio waves	Radio had become used in innumerable ways beyond its better-known areas of telephony, and broadcast television and radio entertainment. Other uses included - emergency services, radar (navigation and weather prediction), sonar, medicine, astronomy, wireless communications, and networking. Radio waves also led researchers to adjacent frequencies such as microwaves, used worldwide for heating and cooking food.

## Scientific community

The scientific community is the group of all interacting scientists. It includes many sub-communities working on particular scientific fields, and within particular institutions; interdisciplinary and cross-institutional activities are also significant.

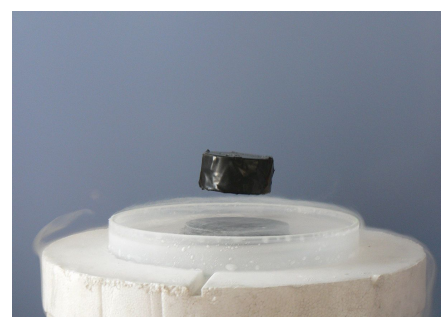
### Branches and fields

Scientific fields are commonly divided into two major groups: natural sciences, which study natural phenomena (including biological life), and social sciences, which study human behavior and societies. These groupings are empirical sciences, which means the knowledge must be based on observable phenomena and capable of being tested for its validity by other researchers working under the same conditions.<sup>[68]</sup> There are also related disciplines that are grouped into interdisciplinary and applied sciences, such as engineering and medicine. Within these categories are specialized scientific fields that can include parts of other scientific disciplines but often possess their own nomenclature and expertise.<sup>[69]</sup>

Mathematics, which is classified as a formal science,<sup>[70][71]</sup> has both similarities and differences with the empirical sciences (the natural and social sciences). It is similar to empirical sciences in that it involves an objective, careful and systematic study of an area of knowledge; it is different because of its method of verifying its knowledge, using *a priori* rather than empirical methods.<sup>[72]</sup> The formal sciences, which also include statistics and logic, are vital to the empirical sciences. Major advances in formal science have often led to major advances in the empirical sciences. The formal sciences are essential in the formation of hypotheses, theories, and laws,<sup>[73]</sup> both in discovering and describing how things work (natural sciences) and how people think and act (social sciences).



Johannes Hevelius and wife Elisabetha making observations, 1673. The Royal Society numbers Hevelius among its first foreign members.



The Meissner effect causes a magnet to levitate above a superconductor

## Institutions

Learned societies for the communication and promotion of scientific thought and experimentation have existed since the Renaissance period.<sup>[74]</sup> The oldest surviving institution is the Italian *Accademia dei Lincei* which was established in 1603.<sup>[75]</sup> The respective National Academies of Science are distinguished institutions that exist in a number of countries, beginning with the British Royal Society in 1660<sup>[76]</sup> and the French *Académie des Sciences* in 1666.<sup>[77]</sup>

International scientific organizations, such as the International Council for Science, have since been formed to promote cooperation between the scientific communities of different nations. Many governments have dedicated agencies to support scientific research. Prominent scientific organizations include, the National Science Foundation in the U.S., the National Scientific and Technical Research Council in Argentina, the academies of science of many nations, CSIRO in Australia, Centre national de la recherche scientifique in France, Max Planck Society and Deutsche Forschungsgemeinschaft in Germany, and in Spain, CSIC.



Louis XIV visiting the Académie des sciences in 1671

## Literature

An enormous range of scientific literature is published.<sup>[78]</sup> Scientific journals communicate and document the results of research carried out in universities and various other research institutions, serving as an archival record of science. The first scientific journals, *Journal des Sçavans* followed by the *Philosophical Transactions*, began publication in 1665. Since that time the total number of active periodicals has steadily increased. As of 1981, one estimate for the number of scientific and technical journals in publication was 11,500.<sup>[79]</sup> The United States National Library of Medicine currently indexes 5,516 journals that contain articles on topics related to the life sciences. Although the journals are in 39 languages, 91 percent of the indexed articles are published in English.<sup>[80]</sup>

Most scientific journals cover a single scientific field and publish the research within that field; the research is normally expressed in the form of a scientific paper. Science has become so pervasive in modern societies that it is generally considered necessary to communicate the achievements, news, and ambitions of scientists to a wider populace.

Science magazines such as *New Scientist*, *Science & Vie*, and *Scientific American* cater to the needs of a much wider readership and provide a non-technical summary of popular areas of research, including notable discoveries and advances in certain fields of research. Science books engage the interest of many more people. Tangentially, the science fiction genre, primarily fantastic in nature, engages the public imagination and transmits the ideas, if not the methods, of science.

Recent efforts to intensify or develop links between science and non-scientific disciplines such as Literature or, more specifically, Poetry, include the *Creative Writing Science* resource developed through the Royal Literary Fund.<sup>[81]</sup>

## Science and society



### Women in science

Science is largely a male-dominated field, with notable exceptions.<sup>[82]</sup> A large majority of male scientists are the ones who have made the discoveries, written the books and thus have written the rules of what to study and how to study it. There is evidence suggesting that this is a product of stereotypes (e.g. science as "manly") as well as self-fulfilling prophecies.<sup>[83][84]</sup> Experiments have shown that parents challenge and explain more to boys than girls, asking them to reflect more deeply and logically.<sup>[85]</sup> Physicist Evelyn Fox Keller argues that science has masculine stereotypes causing ego and competitiveness to obstruct progress, and that these tendencies prevent collaboration and sharing of information.<sup>[86]</sup>

Women have faced a lot of discrimination getting not only credit for their scientific discoveries but also getting opportunities. Both in research and professorship the quantity of females are very limited in comparison to their male counterparts.<sup>[87]</sup> The lack of females in science can be directly associated with the social atmosphere which has always treated science as a more masculine area of study.<sup>[88]</sup> Beginning with boys being pushed more towards academia and girls being confined to the domestic sphere, females have faced both discrimination and difficulty entering into science. Those who are a part of the scientific community find it difficult to break the "glass ceiling" which thus limits how far they can advance within the field. The barrier between work and home has also been an obstacle that women have had to overcome to succeed in the sciences. The achievements of women in science are attributed to their defiance of traditional status of being a laborer within the domestic sphere.<sup>[89]</sup>

Feminist authors and leaders who hail from various educational backgrounds such as Londa Schiebinger, Anne Fausto-Sterling, Bonnie Spanier, and Evelyn Fox Keller<sup>[90]</sup> have published many works interpreting and critiquing science from a feminist perspective. Some criticisms include the gendered metaphors in science, the lack of representation of females in the sciences, how science is used to back up the ideals of patriarchy, and sex/gender dichotomies. Feminist Science Studies as a sub-genre of Women's Studies or Gender Studies are available as areas of study in many universities as a method of activism to promote and encourage awareness of social issues as well as promoting women and intersex individuals to contribute more to the sciences.

Although it has been difficult for women to break into the field of science as credible contributors, many new discoveries have been a result of work by female scientists. Some of the most famous females in the field include Marie Curie, who made discoveries relating to radioactivity, Rosalind Franklin, who worked with x-ray diffraction, Caroline Herschel, who was the first woman to be paid for her scientific work and Jane Goodall, who is currently the world's foremost primatologist.<sup>[91]</sup> These women helped establish a place for women in a heavily male dominated field.<sup>[87]</sup> Most female scientists have only gained fame and authority in the 20th century although there have been advancements in the natural sciences made by women since the early 15th century. Christine de Pizan wrote the first encyclopedia in which she gave credit to these 15th century women for the scientific discoveries of bread making, wool dyeing, grain cultivation and many other day to day inventions.

### Science policy

Science policy is an area of public policy concerned with the policies that affect the conduct of the scientific enterprise, including research funding, often in pursuance of other national policy goals such as technological innovation to promote commercial product development, weapons development, health care and environmental monitoring. Science policy also refers to the act of applying scientific knowledge and consensus to the development of public policies. Science policy thus deals with the entire domain of issues that involve the natural sciences. In



accordance with public policy being concerned about the well-being of its citizens, science policy's goal is to consider how science and technology can best serve the public.

State policy has influenced the funding of public works and science for thousands of years, dating at least from the time of the Mohists, who inspired the study of logic during the period of the Hundred Schools of Thought, and the study of defensive fortifications during the Warring States Period in China. In Great Britain, governmental approval of the Royal Society in the seventeenth century recognized a scientific community which exists to this day. The professionalization of science, begun in the nineteenth century, was partly enabled by the creation of scientific organizations such as the National Academy of Sciences, the Kaiser Wilhelm Institute, and State funding of universities of their respective nations. Public policy can directly affect the funding of capital equipment, intellectual infrastructure for industrial research, by providing tax incentives to those organizations that fund research. Vannevar Bush, director of the office of scientific research and development for the United States government, the forerunner of the National Science Foundation, wrote in July 1945 that "Science is a proper concern of government" <sup>[92]</sup>

Science and technology research is often funded through a competitive process, in which potential research projects are evaluated and only the most promising receive funding. Such processes, which are run by government, corporations or foundations, allocate scarce funds. Total research funding in most developed countries is between 1.5% and 3% of GDP. <sup>[93]</sup> In the OECD, around two-thirds of research and development in scientific and technical fields is carried out by industry, and 20% and 10% respectively by universities and government. The government funding proportion in certain industries is higher, and it dominates research in social science and humanities. Similarly, with some exceptions (e.g. biotechnology) government provides the bulk of the funds for basic scientific research. In commercial research and development, all but the most research-oriented corporations focus more heavily on near-term commercialisation possibilities rather than "blue-sky" ideas or technologies (such as nuclear fusion).

## Media perspectives

The mass media face a number of pressures that can prevent them from accurately depicting competing scientific claims in terms of their credibility within the scientific community as a whole. Determining how much weight to give different sides in a scientific debate may require considerable expertise regarding the matter. <sup>[94]</sup> Few journalists have real scientific knowledge, and even beat reporters who know a great deal about certain scientific issues may be ignorant about other scientific issues that they are suddenly asked to cover. <sup>[95][96]</sup>

## Political usage

Many issues damage the relationship of science to the media and the use of science and scientific arguments by politicians. As a very broad generalisation, many politicians seek certainties and *facts* whilst scientists typically offer probabilities and caveats. However, politicians' ability to be heard in the mass media frequently distorts the scientific understanding by the public. Examples in Britain include the controversy over the MMR inoculation, and the 1988 forced resignation of a Government Minister, Edwina Currie for revealing the high probability that battery farmed eggs were contaminated with *Salmonella*. <sup>[97]</sup>

John Horgan, Chris Mooney, and researchers from the US and Canada have described Scientific Certainty Argumentation Methods (SCAMs), where an organization or think tank makes it their only goal to cast doubt on supported science because it conflicts with political agendas. <sup>[98][99][100][101]</sup> Hank Campbell and microbiologist Alex Berezow have described "feel-good fallacies" used in politics, where politicians frame their positions in a way that makes people feel good about supporting certain policies even when scientific evidence shows there is no need to worry or there is no need for dramatic change on current programs. <sup>[102]</sup>

## Notes

- [1] "... modern science is a discovery as well as an invention. It was a discovery that nature generally acts regularly enough to be described by laws and even by mathematics; and required invention to devise the techniques, abstractions, apparatus, and organization for exhibiting the regularities and securing their law-like descriptions." —p.vii, J. L. Heilbron,(2003, editor-in-chief). *The Oxford Companion to the History of Modern Science*. New York: Oxford University Press. ISBN 0-19-511229-6.
- "science" (<http://www.merriam-webster.com/dictionary/science>). *Merriam-Webster Online Dictionary*. Merriam-Webster, Inc. . Retrieved 2011-10-16. "**3 a:** knowledge or a system of knowledge covering general truths or the operation of general laws especially as obtained and tested through scientific method **b:** such knowledge or such a system of knowledge concerned with the physical world and its phenomena"
- [2] Aristotle, ca. 4th century BCE "[[Nicomachean Ethics (<http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0054:booker=1139b>)] Book VI, and Metaphysics Book I: ". . "In general the sign of knowledge or ignorance is the ability to teach, and for this reason we hold that art rather than experience is scientific knowledge (*epistemē*); for the artists can teach, but the others cannot." — Aristot. *Met.* 1.981b (<http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0052:book=1:section=981b&highlight=artists,others>)
- [3] Isaac Newton's *Philosophiæ Naturalis Principia Mathematica* (1687), for example, is translated "Mathematical Principles of Natural Philosophy", and reflects the then-current use of the words "natural philosophy", akin to "systematic study of nature"
- [4] Oxford English Dictionary
- [5] Needham 1954, p. 150
- [6] See the quotation in Homer (8th c. BCE) *Odyssey* 10.302-3
- [7] "Progress or Return" in *An Introduction to Political Philosophy: Ten Essays* by Leo Strauss. (Expanded version of *Political Philosophy: Six Essays* by Leo Strauss, 1975.) Ed. Hilail Gilden. Detroit: Wayne State UP, 1989.
- [8] Strauss and Cropsey eds. *History of Political Philosophy*, Third edition, p.209.
- [9] "... [A] man knows a thing scientifically when he possesses a conviction arrived at in a certain way, and when the first principles on which that conviction rests are known to him with certainty—for unless he is more certain of his first principles than of the conclusion drawn from them he will only possess the knowledge in question accidentally." — Aristotle, *Nicomachean Ethics* 6 (H. Rackham, ed.) Aristot. *Nic. Eth.* 1139b (<http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0054:booker=1139b>)
- [10] Grant, Edward (2007). *A History of Natural Philosophy: From the Ancient World to the Nineteenth Century*. Cambridge University Press. pp. 62–67. ISBN 978-0-521-68957-1.
- [11] "Galileo and the Birth of Modern Science, by Stephen Hawking, American Heritage's Invention & Technology, Spring 2009, Vol. 24, No. 1, p. 36
- [12] "Galileo Project - Pope Urban VIII Biography" (<http://galileo.rice.edu/gal/urban.html>). .
- [13] Wilson, EO. 1998. *Consilience: The unity of knowledge*. New York. Alfred A. Knopf. p334
- [14] Griffith J. 2011. *What is Science?*. In *The Book of Real Answers to Everything!*, ISBN 9781741290073. <http://www.worldtransformation.com/what-is-science/> accessed November 20, 2012.
- [15] This realization is the topic of intersubjective verifiability, as recounted, for example, by Max Born (1949, 1965) *Natural Philosophy of Cause and Chance* ([http://www.archive.org/stream/naturalphilosoph032159mbp/naturalphilosoph032159mbp\\_djvu.txt](http://www.archive.org/stream/naturalphilosoph032159mbp/naturalphilosoph032159mbp_djvu.txt)), who points out that all knowledge, including natural or social science, is also subjective. Page 162: "Thus it dawned upon me that fundamentally everything is subjective, everything without exception. That was a shock."
- [16] "...[T]he logical empiricists thought that the great aim of science was to discover and establish *generalizations*." —Godfrey-Smith 2003, p. 41
- [17] "Bayesianism tries to understand evidence using probability theory." —Godfrey-Smith 2003, p. 203
- [18] Godfrey-Smith 2003, p. 236
- [19] <http://www.npg.org.uk/collections/search/portrait.php?LinkID=mp02303&rNo=1&role=sit>
- [20] <http://pubs.acs.org/doi/abs/10.1021/ed018p203>
- [21] Godfrey-Smith 2003, p. 20
- [22] Godfrey-Smith 2003, pp. 63–7
- [23] Godfrey-Smith 2003, p. 68
- [24] Godfrey-Smith 2003, p. 69
- [25] Popper called this *Conjecture and Refutation* Godfrey-Smith 2003, pp. 117–8
- [26] Karl Popper: *Objective Knowledge* (1972)
- [27] Newton-Smith, W. H. (1994). *The Rationality of Science*. London: Routledge. p. 30. ISBN 0-7100-0913-5.
- [28] Feyerabend 1993
- [29] Feyerabend, Paul (1987). *Farewell To Reason*. Verso. p. 100. ISBN 0-86091-184-5.
- [30] Godfrey-Smith 2003, p. 151 credits Willard Van Orman Quine (1969) "Epistemology Naturalized" *Ontological Relativity and Other Essays* New York: Columbia University Press, as well as John Dewey, with the basic ideas of naturalism — Naturalized Epistemology, but Godfrey-Smith diverges from Quine's position: according to Godfrey-Smith, "A naturalist can think that science can contribute to *answers* to philosophical questions, without thinking that philosophical questions can be replaced by science questions."
- [31] Brugger, E. Christian (2004). "Casebeer, William D. Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition". *The Review of Metaphysics* 58 (2).



- [32] Popper 1996, p. 4
- [33] Stanovich 2007 pg 119–138
- [34] Stanovich 2007 pg 123
- [35] Fleck, Ludwik (1979). Trenn, Thaddeus J.; Merton, Robert K. eds. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press. ISBN 0-226-25325-2. Claims that before a specific fact "existed", it had to be created as part of a social agreement within a community. Steven Shapin (1980) "A view of scientific thought" *Science* cvii (7 Mar 1980) 1065-66 states "[To Fleck,] facts are invented, not discovered. Moreover, the appearance of scientific facts as discovered things is itself a social construction: a *made* thing."
- [36] Dawkins, Richard; Coyne, Jerry (2005-09-02). "One side can be wrong" (<http://www.guardian.co.uk/science/2005/sep/01/schools.research>). *The Guardian* (London). .
- [37] "Barry Stroud on Scepticism" (<http://philosophybytes.com/2007/12/barry-stroud-on.html>). philosophy bites. 2007-12-16. . Retrieved 2012-02-05.
- [38] Peirce (1877), "The Fixation of Belief", *Popular Science Monthly*, v. 12, pp. 1–15, see §IV on p. 6–7 (<http://books.google.com/books?id=ZKMVAAAAYAAJ&pg=PA6>). Reprinted *Collected Papers* v. 5, paragraphs 358–87 (see 374–6), *Writings* v. 3, pp. 242–57 (see 247–8), *Essential Peirce* v. 1, pp. 109–23 (see 114–15), and elsewhere.
- [39] Peirce (1905), "Issues of Pragmaticism", *The Monist*, v. XV, n. 4, pp. 481–99, see "Character V" on p. 491 (<http://www.archive.org/stream/monistquart15hegeuoft#page/491/mode/1up>). Reprinted in *Collected Papers* v. 5, paragraphs 438–63 (see 451), *Essential Peirce* v. 2, pp. 346–59 (see 353), and elsewhere.
- [40] Peirce (1868), "Some Consequences of Four Incapacities", *Journal of Speculative Philosophy* v. 2, n. 3, pp. 140–57, see p. 141 ([http://books.google.com/books?id=YHkqP2JHJ\\_IC&pg=RA1-PA141](http://books.google.com/books?id=YHkqP2JHJ_IC&pg=RA1-PA141)). Reprinted in *Collected Papers*, v. 5, paragraphs 264–317, *Writings* v. 2, pp. 211–42, *Essential Peirce* v. 1, pp. 28–55, and elsewhere.
- [41] Stanovich 2007 pp 141–147
- [42] "*Pseudoscientific — pretending to be scientific, falsely represented as being scientific*", from the *Oxford American Dictionary*, published by the Oxford English Dictionary; Hansson, Sven Ove (1996). "Defining Pseudoscience", *Philosophia Naturalis*, 33: 169–176, as cited in "Science and Pseudo-science" (<http://plato.stanford.edu/entries/pseudo-science/#NonSciPosSci>) (2008) in Stanford Encyclopedia of Philosophy. The Stanford article states: "Many writers on pseudoscience have emphasized that pseudoscience is non-science posing as science. The foremost modern classic on the subject (Gardner 1957) bears the title *Fads and Fallacies in the Name of Science*. According to Brian Baigrie (1988, 438), "[w]hat is objectionable about these beliefs is that they masquerade as genuinely scientific ones." These and many other authors assume that to be pseudoscientific, an activity or a teaching has to satisfy the following two criteria (Hansson 1996): (1) it is not scientific, and (2) its major proponents try to create the impression that it is scientific".
- For example, Hewitt et al. *Conceptual Physical Science* Addison Wesley; 3 edition (July 18, 2003) ISBN 0-321-05173-4, Bennett et al. *The Cosmic Perspective* 3e Addison Wesley; 3 edition (July 25, 2003) ISBN 0-8053-8738-2; See also, e.g., Gauch HG Jr. *Scientific Method in Practice* (2003).
  - A 2006 National Science Foundation report on Science and engineering indicators quoted Michael Shermer's (1997) definition of pseudoscience: "'claims presented so that they appear [to be] scientific even though they lack supporting evidence and plausibility'"(p. 33). In contrast, science is "a set of methods designed to describe and interpret observed and inferred phenomena, past or present, and aimed at building a testable body of knowledge open to rejection or confirmation"(p. 17). Shermer M. (1997). *Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time*. New York: W. H. Freeman and Company. ISBN 0-7167-3090-1. as cited by National Science Board. National Science Foundation, Division of Science Resources Statistics (2006). "Science and Technology: Public Attitudes and Understanding" (<http://www.nsf.gov/statistics/seind06/c7/c7s2.htm>). *Science and engineering indicators 2006*. .
  - "A pretended or spurious science; a collection of related beliefs about the world mistakenly regarded as being based on scientific method or as having the status that scientific truths now have," from the *Oxford English Dictionary*, second edition 1989.
- [43] Cargo Cult Science ([http://neurotheory.columbia.edu/~ken/cargo\\_cult.html](http://neurotheory.columbia.edu/~ken/cargo_cult.html)) by Feynman, Richard. Retrieved 2011-07-21.
- [44] "Coping with fraud" (<http://web.archive.org/web/20070928151119/http://www.publicationethics.org.uk/reports/1999/1999pdf3.pdf>) (PDF). *The COPE Report 1999*: 11–18. Archived from the original (<http://www.publicationethics.org.uk/reports/1999/1999pdf3.pdf>) on 2007-09-28. . Retrieved 2011-07-21. "It is 10 years, to the month, since Stephen Lock ... Reproduced with kind permission of the Editor, The Lancet."
- [45] In mathematics, Plato's *Meno* demonstrates that it is possible to know logical propositions, such as the Pythagorean theorem, and even to prove them, as cited by Crease 2009, pp. 35–41
- [46] Ziman cites Polanyi 1958 chapter 12, as referenced in Ziman 1978, p. 44
- [47] Ziman 1978, pp. 46–47
- [48] Ziman 1978, p. 46
- [49] Ziman 1978, p. 104.
- [50] di Francia 1976, p. 13: "The amazing point is that for the first time since the discovery of mathematics, a method has been introduced, the results of which have an intersubjective value!" (*Author's punctuation*)
- [51] Wilson, Edward (1999). *Consilience: The Unity of Knowledge*. New York: Vintage. ISBN 0-679-76867-X
- [52] di Francia 1976, pp. 4–5: "One learns in a laboratory; one learns how to make experiments only by experimenting, and one learns how to work with his hands only by using them. The first and fundamental form of experimentation in physics is to teach young people to work with their hands. Then they should be taken into a laboratory and taught to work with measuring instruments — each student carrying out real experiments in physics. This form of teaching is indispensable and cannot be read in a book."

- [53] Fara 2009, p. 204: "Whatever their discipline, scientists claimed to share a common scientific method that ... distinguished them from non-scientists."
- [54] Nola & Irzik 2005, p. 208
- [55] Nola & Irzik 2005, pp. 199–201
- [56] van Gelder, Tim (1999). "'Heads I win, tails you lose': A Foray Into the Psychology of Philosophy" (<http://web.archive.org/web/20080409054240/http://www.philosophy.unimelb.edu.au/tgelder/papers/HeadsIWin.pdf>) (PDF). University of Melbourne. Archived from the original (<http://www.philosophy.unimelb.edu.au/tgelder/papers/HeadsIWin.pdf>) on 2008-04-09. . Retrieved 2008-03-28.
- [57] Pease, Craig (September 6, 2006). "Chapter 23. Deliberate bias: Conflict creates bad science" ([http://web.archive.org/web/20100619154617/http://law-and-science.net/Science4BLJ/Scientific\\_Method/Deliberate.bias/Text.htm](http://web.archive.org/web/20100619154617/http://law-and-science.net/Science4BLJ/Scientific_Method/Deliberate.bias/Text.htm)). *Science for Business, Law and Journalism*. Vermont Law School. Archived from the original on 19 June 2010. .
- [58] Shatz, David (2004). *Peer Review: A Critical Inquiry*. Rowman & Littlefield. ISBN 0-7425-1434-X. OCLC 54989960.
- [59] Krimsky, Sheldon (2003). *Science in the Private Interest: Has the Lure of Profits Corrupted the Virtue of Biomedical Research*. Rowman & Littlefield. ISBN 0-7425-1479-X. OCLC 185926306.
- [60] Bulger, Ruth Ellen; Heitman, Elizabeth; Reiser, Stanley Joel (2002). *The Ethical Dimensions of the Biological and Health Sciences* (2nd ed.). Cambridge University Press. ISBN 0-521-00886-7. OCLC 47791316.
- [61] Backer, Patricia Ryaby (October 29, 2004). "What is the scientific method?" ([http://www.engr.sjsu.edu/pabacker/scientific\\_method.htm](http://www.engr.sjsu.edu/pabacker/scientific_method.htm)). San Jose State University. . Retrieved 2008-03-28.
- [62] Graduate Education for Computational Science and Engineering (<http://www.siam.org/students/resources/report.php>), SIAM Working Group on CSE Education. Retrieved 2008-04-27.
- [63] Bunge, Mario Augusto (1998). *Philosophy of Science: From Problem to Theory*. Transaction Publishers. p. 24. ISBN 0-7658-0413-1.
- [64] "To Live at All Is Miracle Enough — Richard Dawkins" (<http://richarddawkins.net/articles/91>). RichardDawkins.net. 2006-05-10. . Retrieved 2012-02-05.
- [65] Stanovich 2007, pp. 106–110
- [66] Simonton, Dean Keith. "After Einstein: Scientific genius is extinct". *Nature* **493** (7434): 602–602. doi:10.1038/493602a.
- [67] Evicting Einstein ([http://science.nasa.gov/science-news/science-at-nasa/2004/26mar\\_einstein](http://science.nasa.gov/science-news/science-at-nasa/2004/26mar_einstein)), March 26, 2004, NASA. "*Both [relativity and quantum mechanics] are extremely successful. The Global Positioning System (GPS), for instance, wouldn't be possible without the theory of relativity. Computers, telecommunications, and the Internet, meanwhile, are spin-offs of quantum mechanics.*"
- [68] Popper 2002, p. 20
- [69] See: Editorial Staff (March 7, 2008). "Scientific Method: Relationships among Scientific Paradigms" ([http://www.seedmagazine.com/news/2007/03/scientific\\_method\\_relationship.php](http://www.seedmagazine.com/news/2007/03/scientific_method_relationship.php)). Seed magazine. . Retrieved 2007-09-12.
- [70] "Marcus Tomalin (2006) "Linguistics and the Formal Sciences"" ([http://www.cambridge.org/gb/knowledge/isbn/item1173067/?site\\_locale=en\\_GB](http://www.cambridge.org/gb/knowledge/isbn/item1173067/?site_locale=en_GB)). Cambridge.org. doi:10.2277/0521854814. . Retrieved 2012-02-05.
- [71] Benedikt Löwe (2002) "The Formal Sciences: Their Scope, Their Foundations, and Their Unity" (<http://www.jstor.org/pss/20117289>)
- [72] Popper 2002, pp. 10–11
- [73] Popper 2002, pp. 79–82
- [74] Parrott, Jim (August 9, 2007). "Chronicle for Societies Founded from 1323 to 1599" (<http://www.scholarly-societies.org/1599andearlier.html>). Scholarly Societies Project. . Retrieved 2007-09-11.
- [75] "Accademia Nazionale dei Lincei" (<http://positivamente.lincoi.it/>) (in Italian). 2006. . Retrieved 2007-09-11.
- [76] "History of the Royal Society" (<http://royalsociety.org/about-us/history/>). The Royal Society. . Retrieved 2011-10-16.
- [77] Meynell, G.G.. "The French Academy of Sciences, 1666–91: A reassessment of the French Académie royale des sciences under Colbert (1666–83) and Louvois (1683–91)" ([http://www.haven.u-net.com/6text\\_7B2.htm#Appendix 2](http://www.haven.u-net.com/6text_7B2.htm#Appendix%202)). . Retrieved 2011-10-13.
- [78] Ziman, J.M. (1980). "The proliferation of scientific literature: a natural process". *Science* **208** (4442): 369–371. doi:10.1126/science.7367863. PMID 7367863.
- [79] Subramanyam, Krishna; Subramanyam, Bhadriraju (1981). *Scientific and Technical Information Resources*. CRC Press. ISBN 0-8247-8297-6. OCLC 232950234.
- [80] "MEDLINE Fact Sheet" (<http://www.nlm.nih.gov/pubs/factsheets/medline.html>). Washington DC: United States National Library of Medicine. . Retrieved 2011-10-15.
- [81] Petrucci, Mario. "Creative Writing <-> Science" (<http://writeideas.org.uk/creativescience/index.htm>). . Retrieved 2008-04-27.
- [82] Women in science have included:
- Hypatia (c.350–415 CE), of the Library of Alexandria.
  - Trotula of Salerno, a physician c.1060 CE.
  - Caroline Herschel one of the first professional astronomers of the 18th and 19th c.
  - Christine Ladd-Franklin, a doctoral student of C. S. Peirce, who published Wittgenstein's proposition 5.101 in her dissertation, 40 years before Wittgenstein's publication of *Tractatus Logico-Philosophicus*.
  - Henrietta Leavitt, a professional human computer and astronomer, who first published the significant relationship between the luminosity of Cepheid variable stars and their distance from Earth. This allowed Hubble to make the discovery of the expanding universe, which led to the Big Bang theory.
  - Emmy Noether, who proved the conservation of energy and other constants of motion in 1915.
  - Nina Byers notes that after 1976, women in science became much more prevalent in science, than the exceptions

- [83] Summers, L. H. (2005). Remarks at NBER Conference on Diversifying the Science & Engineering Workforce. The office of the President. Harvard University.
- [84] Nosek, B.A., et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *PNAS*, June 30, 2009, 106, 10593–10597.
- [85] Crowley, K. Callanan, M.A., Tenenbaum, H. R., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. *Psychological Science*, 258–261.
- [86] Reflections on Gender and Science. Yale University Press, 1985.
- [87] Londa Schiebinger, *Has Feminism Changed Science* (Cambridge: Harvard University Press, 1999)
- [88] Londa Schiebinger, *Has Feminism Changed Science* (Cambridge: Harvard University Press, 1999).
- [89] Bonnie Spanier, *From Molecules to Brains, Normal Science Supports Sexist Beliefs About Differences*, *The Gender and Science Reader* (New York: Routledge 2001)
- [90] *The Gender and Science Reader*
- [91] Sarah Zielinski, *Ten Historic Female Scientists You Should Know*, web.<http://www.smithsonianmag.com/science-nature/Ten-Historic-Female-Scientists-You-Should-Know.html>.
- [92] "Vannevar Bush (July 1945), "Science, the Endless Frontier"" (<http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>). Nsf.gov. . Retrieved 2012-02-05.
- [93] "Main Science and Technology Indicators - 2008-1" (<http://www.webcitation.org/5tamBWVvW>) (pdf). OECD. . Retrieved 20 April 2012. 50.8 KB
- [94] Dickson, David (October 11, 2004). "Science journalism must keep a critical edge" (<http://web.archive.org/web/20100621053624/http://www.scidev.net/en/editorials/science-journalism-must-keep-a-critical-edge.html>). Science and Development Network. Archived from the original on 21 June 2010. .
- [95] Mooney, Chris (Nov/Dec 2004). *Blinded By Science, How 'Balanced' Coverage Lets the Scientific Fringe Hijack Reality* (<http://blogs.discovermagazine.com/intersection/2010/01/15/blinded-by-science-how-balanced-coverage-lets-the-scientific-fringe-hijack-reality/>). **43**. *Columbia Journalism Review*. . Retrieved 2008-02-20.
- [96] McIlwaine, S.; Nguyen, D. A. (2005). "Are Journalism Students Equipped to Write About Science?" (<http://espace.library.uq.edu.au/view/UQ:8064>). *Australian Studies in Journalism* **14**: 41–60. . Retrieved 2008-02-20.
- [97] "1988: Egg industry fury over salmonella claim" ([http://news.bbc.co.uk/onthisday/hi/dates/stories/december/3/newsid\\_2519000/2519451.stm](http://news.bbc.co.uk/onthisday/hi/dates/stories/december/3/newsid_2519000/2519451.stm)), "On This Day," BBC News, December 3, 1988.
- [98] "Original "Doubt is our product..." memo" (<http://legacy.library.ucsf.edu/tid/wjh13f00>). University of California, San Francisco. 21 August 1969. . Retrieved 3 October 2012. The memo reads "Doubt is our product since it is the best means of competing with the 'body of fact' that exists in the mind of the general public. It is also the means of establishing a controversy."
- [99] "'THE REPUBLICAN WAR ON SCIENCE,' BY CHRIS MOONEY", *Political Science*, Review by JOHN HORGAN, Published: December 18 2005 (<http://www.nytimes.com/2005/12/18/books/review/18horgan.html?pagewanted=all>)
- [100] Mooney, Chris (2005). *The Republican War on Science*. Basic Books. ISBN 0-465-04676-2.
- [101] William R. Freudenburg, Robert Gramling, Debra J. Davidson (2008) (<http://onlinelibrary.wiley.com/doi/10.1111/j.1475-682X.2008.00219.x/abstract>) "Scientific Certainty Argumentation Methods (SCAMs): Science and the politics of doubt". *Sociological Inquiry*. Vol. **78**, No. 1. 2–38
- [102] Hank Campbell, Alex Berezow., *Science Left Behind : Feel-good Fallacies and the Rise of the Anti-Scientific Left* (1st ed.). New York: PublicAffairs. ISBN 9781610391641.

## References

- Crease, Robert P. (2009). *The Great Equations*. New York: W.W. Norton. p. 317. ISBN 978-0-393-06204-5
- Crease, Robert P. (2011). *World in the Balance: the historic quest for an absolute system of measurement*. New York: W.W. Norton. p. 317. ISBN 978-0-393-07298-3.
- di Francia, Giuliano Toraldo (1976). *The Investigation of the Physical World*. Cambridge: Cambridge University Press. ISBN 0-521-29925-X Originally published in Italian as *L'Indagine del Mondo Fisico* by Giulio Einaudi editore 1976; first published in English by Cambridge University Press 1981.
- Fara, Patricia (2009). *Science : a four thousand year history*. Oxford: Oxford University Press. p. 408. ISBN 978-0-19-922689-4.
- Feyerabend, Paul (1993). *Against Method* (3rd ed.). London: Verso. ISBN 0-86091-646-4.
- Feyerabend, Paul (2005). *Science, history of the philosophy*, as cited in Honderich, Ted (2005). *The Oxford companion to philosophy*. Oxford Oxfordshire: Oxford University Press. ISBN 0-19-926479-1. OCLC 173262485.
- Godfrey-Smith, Peter (2003). *Theory and Reality*. Chicago 60637: University of Chicago. p. 272. ISBN 0-226-30062-5

- Feynman, R.P. (1999). *The Pleasure of Finding Things Out: The Best Short Works of Richard P. Feynman*. Perseus Books Group. ISBN 0-465-02395-9. OCLC 181597764.
- Needham, Joseph (1954). *Science and Civilisation in China: Introductory Orientations. 1*. Cambridge University Press
- Nola, Robert; Irzik, Gürol (2005). *Philosophy, science, education and culture* (<http://books.google.com/books?id=HRqDnikufC0C>). Science & technology education library. **28**. Springer. ISBN 1-4020-3769-4.
- Papineau, David. (2005). *Science, problems of the philosophy of.*, as cited in Honderich, Ted (2005). *The Oxford companion to philosophy*. Oxford Oxfordshire: Oxford University Press. ISBN 0-19-926479-1. OCLC 173262485.
- Parkin, D. (1991). Philip M. Peek. ed. *African Divination Systems: Ways of Knowing*. Indianapolis, IN: Indiana University Press..
- Polanyi, Michael (1958). *Personal Knowledge: Towards a Post-Critical Philosophy*. University of Chicago Press. ISBN 0-226-67288-3
- Popper, Karl Raimund (1996) [1984]. *In search of a better world: lectures and essays from thirty years* (<http://books.google.com/books?id=L33XSZE77OkC>). New York, NY: Routledge. ISBN 0-415-13548-6.
- Popper, Karl R. (2002) [1959]. *The Logic of Scientific Discovery*. New York, NY: Routledge Classics. ISBN 0-415-27844-9. OCLC 59377149.
- Stanovich, Keith E. (2007). *How to Think Straight About Psychology*. Boston: Pearson Education. ISBN 978-0-205-68590-5.
- Ziman, John (1978). *Reliable knowledge: An exploration of the grounds for belief in science*. Cambridge: Cambridge University Press. p. 197. ISBN 0-521-22087-4

## Further reading

- Augros, Robert M., Stanciu, George N., "The New Story of Science: mind and the universe", Lake Bluff, Ill.: Regnery Gateway, c1984. ISBN 0-89526-833-7
- Becker, Ernest (1968). *The structure of evil; an essay on the unification of the science of man*. New York: G. Braziller.
- Cole, K. C., *Things your teacher never told you about science: Nine shocking revelations* Newsday, Long Island, New York, March 23, 1986, pg 21+
- Feynman, Richard "Cargo Cult Science" (<http://calteches.library.caltech.edu/51/02/CargoCult.pdf>)
- Gaukroger, Stephen (2006). *The Emergence of a Scientific Culture: Science and the Shaping of Modernity 1210–1685*. Oxford: Oxford University Press. ISBN 0-19-929644-8.
- Gopnik, Alison, "Finding Our Inner Scientist" (<http://www.amacad.org/publications/winter2004/gopnik.pdf>), Daedalus, Winter 2004.
- Krige, John, and Dominique Pestre, eds., *Science in the Twentieth Century*, Routledge 2003, ISBN 0-415-28606-9
- Levin, Yuval (2008). *Imagining the Future: Science and American Democracy*. New York, Encounter Books. ISBN 1-59403-209-2
- Kuhn, Thomas, *The Structure of Scientific Revolutions*, 1962.
- William F., McComas (1998). "The principal elements of the nature of science: Dispelling the myths" ([http://earthweb.ess.washington.edu/roe/Knowability\\_590/Week2/Myths of Science.pdf](http://earthweb.ess.washington.edu/roe/Knowability_590/Week2/Myths%20of%20Science.pdf)). In McComas, William F.. *The nature of science in science education: rationales and strategies*. Springer. ISBN 978-0-7923-6168-8
- Obler, Paul C.; Estrin, Herman A. (1962). *The New Scientist: Essays on the Methods and Values of Modern Science*. Anchor Books, Doubleday.
- Russell, Bertrand (1985) [1952]. *The Impact of Science on Society*. London: Unwin. ISBN 0-04-300090-8.
- Rutherford, F. James; Ahlgren, Andrew (1990). *Science for all Americans*. New York, NY: American Association for the Advancement of Science, Oxford University Press. ISBN 0-19-506771-1.

- Thurs, Daniel Patrick (2007). *Science Talk: Changing Notions of Science in American Popular Culture*. New Brunswick, NJ: Rutgers University Press. pp. 22–52. ISBN 978-0-8135-4073-3.

## External links

### Publications

- "GCSE Science textbook". Wikibooks.org

### News

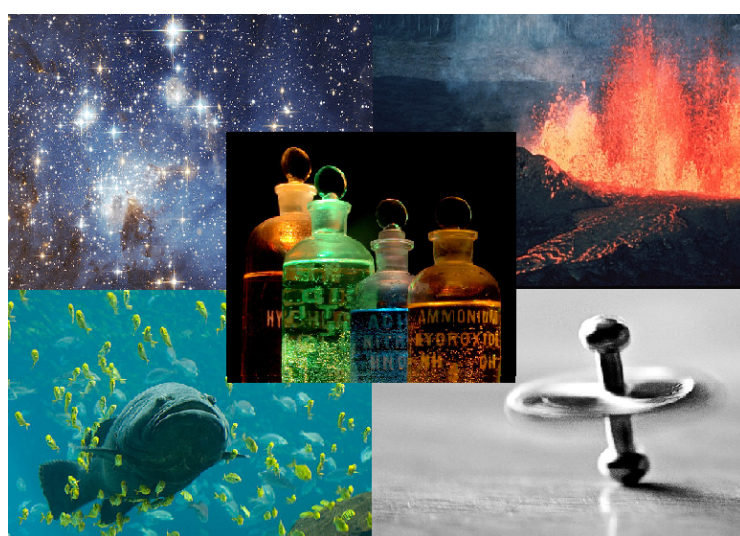
- Nature News (<http://www.nature.com/news/>). Science news by the journal *Nature*
- New Scientist (<http://www.newscientist.com/>). An weekly magazine published by Reed Business Information
- ScienceDaily (<http://www.sciencedaily.com/>)
- Science Newsline (<http://www.sciencenewsline.com/>)
- Scienica (<http://scienica.com/>)
- Discover Magazine (<http://www.discovermagazine.com/>)
- Irish Science News (<http://www.science.ie/>) from Discover Science & Engineering
- Science Stage (<http://sciencestage.com/>) Scientific Videoportal and Community

### Resources

- Euroscience (<http://www.euroscience.org/>):
  - "ESOF: Euroscience Open Forum" (<http://web.archive.org/web/20100610204548/http://www.euroscience.org/esof.html>). Archived from the original on 10 June 2010.
- Science Development in the *Latin American docta* ([http://www.en.argentina.ar/\\_en/science-and-education/](http://www.en.argentina.ar/_en/science-and-education/))
- Classification of the Sciences (<http://xtf.lib.virginia.edu/xtf/view?docId=DicHist/uvaBook/tei/DicHist1.xml;chunk.id=dv1-57;toc.depth=1;toc.id=dv1-57;brand=default>) in *Dictionary of the History of Ideas*. (Dictionary's new electronic format is badly botched, entries after "Design" are inaccessible. *Internet Archive* old version (<http://web.archive.org/web/20080619205103/http://etext.lib.virginia.edu/cgi-local/DHI/dhi.cgi?id=dv1-57>)).
- "Nature of Science" (<http://evolution.berkeley.edu/evosite/nature/index.shtml>) University of California Museum of Paleontology
- United States Science Initiative (<http://www.science.gov/>) Selected science information provided by US Government agencies, including research & development results
- How science works (<http://undsci.berkeley.edu/index.php>) University of California Museum of Paleontology

# Natural science

The **natural sciences** are those branches of science that seek to elucidate the rules that govern the natural world through scientific methods.<sup>[1]</sup> The term "natural science" is used to distinguish the subject from the social sciences, which apply the scientific method to study human behavior and social patterns; the humanities, which use a critical or analytical approach to study the human condition; and the formal sciences such as mathematics and logic, which use an a priori, as opposed to factual methodology to study formal systems.



The natural sciences seek to understand how the world and universe around us works. There are five major branches: Chemistry (center), astronomy, earth science, physics, and biology (clockwise from top-left).

## Overview

There are five branches of natural science: astronomy, biology, chemistry, the Earth sciences and physics.<sup>[2][3]</sup> This distinguishes sciences that cover inquiry into the world of nature from human sciences such as anthropology, sociology and linguistics, and from formal sciences such as mathematics and logic.<sup>[2]</sup> Despite their differences, these sciences sometimes overlap. For example, the social sciences and biology both study human beings as organisms while mathematics is used regularly in all the natural sciences.<sup>[2]</sup>

Alongside its traditional usage, natural science may encompass natural history, which emerged in the 16th century and focused on the description and classification of plants, animals, minerals and other natural objects.<sup>[4]</sup> Today, natural history refers to observational descriptions of the natural world aimed at popular audiences rather than an academic ones.<sup>[5]</sup> The natural sciences are sometimes referred to colloquially as hard science, or fields seen as relying on experimental, quantifiable data or the scientific method and focusing on accuracy and objectivity.<sup>[6]</sup> These usually include physics, chemistry and biology.<sup>[6]</sup> By contrast, soft science is used as a pejorative term to describe fields more reliant on qualitative research, including the social sciences.<sup>[6]</sup>

## History

Some scholars trace the origins of natural science as far back as pre-literate human societies, where understanding the natural world was necessary for survival.<sup>[7]</sup> People observed and built up knowledge about the behavior of animals and the usefulness of plants as food and medicine, which was passed down from generation to generation.<sup>[7]</sup> These primitive understandings gave way to more formalized inquiry around 3,500 to 3,000 B.C. in Mesopotamian and Ancient Egyptian cultures, which produced the first known written evidence of natural philosophy, the precursor of natural science.<sup>[8]</sup> While the writings show an interest in astronomy, mathematics and other aspects of the physical world, the ultimate aim of inquiry about nature's workings was in all cases religious or mythological, not scientific.<sup>[9]</sup> A tradition of scientific inquiry also emerged in Ancient China, where Taoist alchemists and philosophers experimented with elixirs to extend life and cure ailments.<sup>[10]</sup> They focused on the yin and yang, or contrasting elements in nature; the yin was associated with femininity and coldness, while yang was associated with masculinity and warmth.<sup>[11]</sup> The five phases – fire, earth, metal, wood and water – described a cycle of transformations in nature. Water turned into wood, which turned into fire when it burned. The ashes left by fire were earth.<sup>[12]</sup> Using

these principles, Chinese philosophers and doctors explored human anatomy, characterizing organs as predominantly yin or yang; they understood the relationship between the pulse, the heart and the flow of blood in the body centuries before it became accepted in the West.<sup>[13]</sup>

Little evidence survives of how Ancient Indian cultures around the Indus River understood nature, but some of their perspectives may be reflected in the Vedas, a set of sacred Hindu texts.<sup>[13]</sup> They reveal a conception of the universe as ever-expanding and constantly being recycled and reformed.<sup>[13]</sup> Surgeons in the Ayurvedic tradition saw health and illness as a combination of three humors: wind, bile and phlegm.<sup>[13]</sup> A healthy life was the result of a balance between these humors.<sup>[13]</sup> In Ayurvedic thought, the body consisted of five elements: earth, water, fire, wind and empty space.<sup>[13]</sup> Ayurvedic surgeons performed complex surgeries and developed a detailed understanding of human anatomy.<sup>[13]</sup>

Pre-Socratic philosophers in Ancient Greek culture brought natural philosophy a step closer to direct inquiry about cause and effect in nature between 600 and 400 B.C., although an element of magic and mythology remained.<sup>[14]</sup> Natural phenomena such as earthquakes and eclipses were explained increasingly in the context of nature itself instead of being attributed to angry gods.<sup>[14]</sup> Thales of Miletus, an early philosopher who lived from 625 to 546 B.C., explained earthquakes by theorizing that the world floated on water and that water was the fundamental element in nature.<sup>[15]</sup> In the fifth century B.C., Leucippus was an early exponent of atomism, the idea that the world is made up of fundamental indivisible particles.<sup>[16]</sup> Pythagoras applied Greek innovations in mathematics to astronomy, and suggested that the earth was spherical.<sup>[16]</sup>

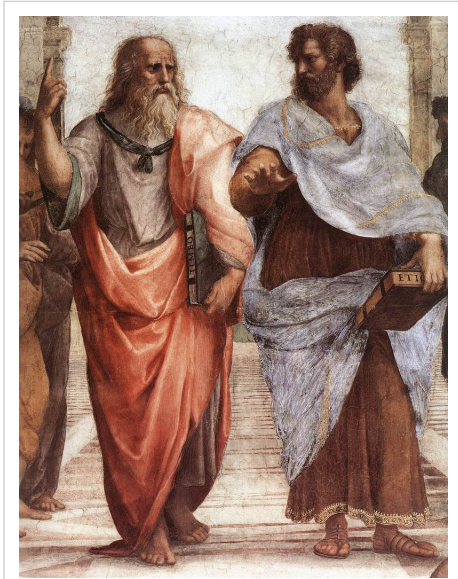
### **Aristotelian natural philosophy (400 B.C.–1100 A.D.)**

Later Socratic and Platonic thought focused on ethics, morals and art and did not attempt an investigation of the physical world; Plato criticized pre-Socratic thinkers as materialists and anti-religionists.<sup>[17]</sup> Aristotle, however, a student of Plato who lived from 384 to 322 B.C., paid closer attention to the natural world in his philosophy.<sup>[18]</sup> In his *History of Animals*, he described the inner workings of 110 species, including the stingray, catfish and bee.<sup>[19]</sup> He investigated chick embryos by breaking open eggs and observing them at various stages of development.<sup>[20]</sup> Aristotle's works were influential through the 19th century, and he is considered by some scholars to be the father of biology.<sup>[21]</sup> He also presented philosophies about physics, nature and astronomy using inductive reasoning in his works *Physics* and *Meteorology*.<sup>[22]</sup>



While Aristotle considered natural philosophy more seriously than his predecessors, he approached it as a theoretical branch of science.<sup>[23]</sup> Still, inspired by his work, Ancient Roman philosophers of the early first century A.D., including Lucretius, Seneca and Pliny the Elder, wrote treatises that dealt with the rules of the natural world in varying degrees of depth.<sup>[24]</sup> Many Ancient Roman Neoplatonists of the third to the sixth centuries A.D. also adapted Aristotle's teachings on the physical world to a philosophy that emphasized spiritualism.<sup>[25]</sup> Early medieval philosophers including Macrobius, Calcidius and Martianus Capella also examined the physical world, largely from a cosmological and cosmographical perspective, putting forth theories on the arrangement of celestial bodies and the heavens, which were posited as being composed of aether.<sup>[26]</sup>

Aristotle's works on natural philosophy continued to be translated and studied amid the rise of the Byzantine Empire and Islam in the Middle East.<sup>[27]</sup> A revival in mathematics and science took place during the time of the Abbasid Caliphate from the ninth century onward, when Muslim scholars expanded upon Greek and Indian natural philosophy.<sup>[28]</sup> The words *alcohol*, *algebra* and *zenith* all have Arabic roots.<sup>[29]</sup>



Plato (left) and Aristotle in a 1509 painting by Raphael. Plato rejected inquiry into natural philosophy as against religion, while his student, Aristotle, created a body of work on the natural world that influenced generations of scholars.

### Medieval natural philosophy (1100–1600)

Aristotle's works and other Greek natural philosophy did not reach the West until about the middle of the 12th century, when works were translated from Greek and Arabic into Latin.<sup>[30]</sup> The development of European civilization later in the Middle Ages brought with it further advances in natural philosophy.<sup>[31]</sup> European inventions such as the horseshoe, horse collar and crop rotation allowed for rapid population growth, eventually giving way to urbanization and the foundation of schools connected to monasteries and cathedrals in modern-day France and England.<sup>[32]</sup> Aided by the schools, an approach to Christian theology developed that sought to answer questions about nature and other subjects using logic.<sup>[33]</sup> This approach, however, was seen by some detractors as heresy.<sup>[33]</sup> By the 12th century, Western European scholars and philosophers came into contact with a body of knowledge of which they had previously been ignorant: a large corpus of works in Greek and Arabic that were preserved by Islamic scholars.<sup>[34]</sup> Through translation into Latin, Western Europe was introduced to Aristotle and his natural philosophy.<sup>[34]</sup> These works were taught at new universities in Paris and Oxford by the early 13th century, although the practice was frowned upon by the Catholic church.<sup>[35]</sup> A 1210 decree from the Synod of Paris ordered that "no lectures are to be held in Paris either publicly or privately using Aristotle's books on natural philosophy or the commentaries, and we forbid all this under pain of excommunication."<sup>[35]</sup>

In the late Middle Ages, Spanish philosopher Dominicus Gundissalinus translated a treatise by the earlier Arab scholar Al-Farabi called *On the Sciences* into Latin, calling the study of the mechanics of nature *scientia naturalis*, or natural science.<sup>[36]</sup> Gundissalinus also proposed his own classification of the natural sciences in his 1150 work *On the Division of Philosophy*.<sup>[36]</sup> This was the first detailed classification of the sciences based on Greek and Arab philosophy to reach Western Europe.<sup>[36]</sup> Gundissalinus defined natural science as "the science considering only things unabstracted and with motion," as opposed to mathematics and sciences that rely on mathematics.<sup>[37]</sup> Following Al-Farabi, he then separated the sciences into eight parts, including physics, cosmology, meteorology, minerals science and plant and animal science.<sup>[37]</sup>



Later philosophers made their own classifications of the natural sciences. Robert Kilwardby wrote *On the Order of the Sciences* in the 13th century that classed medicine as a mechanical science, along with agriculture, hunting and theater while defining natural science as the science that deals with bodies in motion.<sup>[38]</sup> Roger Bacon, an English friar and philosopher, wrote that natural science dealt with "a principle of motion and rest, as in the parts of the elements of fire, air, earth and water, and in all inanimate things made from them."<sup>[39]</sup> These sciences also covered plants, animals and celestial bodies.<sup>[39]</sup> Later in the 13th century, Catholic priest and theologian Thomas Aquinas defined natural science as dealing with "mobile beings" and "things which depend on matter not only for their existence, but also for their definition."<sup>[40]</sup> There was wide agreement among scholars in medieval times that natural science was about bodies in motion, although there was division about the inclusion of fields including medicine, music and perspective.<sup>[41]</sup> Philosophers pondered questions including the existence of a vacuum, whether motion could produce heat, the colors of rainbows, the motion of the earth, whether elemental chemicals exist and where in the atmosphere rain is formed.<sup>[42]</sup>

In the centuries up through the end of the Middle Ages, natural science was often mingled with philosophies about magic and the occult.<sup>[43]</sup> Natural philosophy appeared in a wide range of forms, from treatises to encyclopedias to commentaries on Aristotle.<sup>[44]</sup> The interaction between natural philosophy and Christianity was complex during this period; some early theologians, including Tatian and Eusebius, considered natural philosophy an outcropping of pagan Greek science and were suspicious of it.<sup>[45]</sup> Although some later Christian philosophers, including Aquinas, came to see natural science as a means of interpreting scripture, this suspicion persisted until the 12th and 13th centuries.<sup>[46]</sup> The Condemnation of 1277, which forbade setting philosophy on a level equal with theology and the debate of religious constructs in a scientific context, showed the persistence with which Catholic leaders resisted the development of natural philosophy even from a theological perspective.<sup>[47]</sup> Aquinas and Albertus Magnus, another Catholic theologian of the era, sought to distance theology from science in their works.<sup>[48]</sup> "I don't see what one's interpretation of Aristotle has to do with the teaching of the faith," he wrote in 1271.<sup>[49]</sup>

### **Newton and the scientific revolution (1600–1800)**

By the 16th and 17th centuries, natural philosophy underwent an evolution beyond commentary on Aristotle as more early Greek philosophy was uncovered and translated.<sup>[50]</sup> The invention of the printing press in the 1400s, the invention of the microscope and telescope, and the Protestant Reformation fundamentally altered the social context in which scientific inquiry evolved in the West.<sup>[50]</sup> Christopher Columbus's discovery of a new world changed perceptions about the physical makeup of the world, while observations by Copernicus, Tyco Brahe and Galileo brought a more accurate picture of the solar system as heliocentric and proved many of Aristotle's theories about the heavenly bodies false.<sup>[51]</sup> A number of 17th-century philosophers, including Thomas Hobbes, John Locke and Francis Bacon made a break from the past by rejecting Aristotle and his medieval followers outright, calling their approach to natural philosophy as superficial.<sup>[52]</sup>

The titles of Galileo's work *Two New Sciences* and Johannes Kepler's *New Astronomy* underscored the atmosphere of change that took hold in the 17th century as Aristotle was dismissed in favor of novel methods of inquiry into the natural world.<sup>[53]</sup> Bacon was instrumental in popularizing this change; he argued that people should use the arts and sciences to gain dominion over nature.<sup>[54]</sup> To achieve this, he wrote that "human life [must] be endowed with new discoveries and powers."<sup>[55]</sup> He defined natural philosophy as "the knowledge of Causes and secret motions of things; and enlarging the bounds of Human Empire, to the effecting of all things possible."<sup>[53]</sup> Bacon proposed scientific inquiry supported by the state and fed by the collaborative research of scientists, a vision that was unprecedented in its scope, ambition and form at the time.<sup>[55]</sup> Natural philosophers came to view nature increasingly as a mechanism that could be taken apart and understood, much like a complex clock.<sup>[56]</sup> Natural philosophers including Isaac Newton, Evangelista Torricelli and Francesco Redi conducted experiments focusing on the flow of water, measuring atmospheric pressure using a barometer and disproving spontaneous generation.<sup>[57]</sup> Scientific societies and scientific journals emerged and were spread widely through the printing press, touching off the scientific revolution.<sup>[58]</sup> Newton in 1687 published his *The Mathematical Principles of Natural Philosophy*, or

*Principia Mathematica*, which set the groundwork for physical laws that remained current until the 19th century.<sup>[59]</sup>

Some modern scholars, including Andrew Cunningham, Perry Williams and Floris Cohen, argue that natural philosophy is not properly called a science, and that genuine scientific inquiry began only with the scientific revolution.<sup>[60]</sup> According to Cohen, "the emancipation of science from an overarching entity called 'natural philosophy' is one defining characteristic of the Scientific Revolution."<sup>[60]</sup> Other historians of science, including Edward Grant, contend that the scientific revolution that blossomed in the 17th, 18th and 19th centuries occurred when principles learned in the exact sciences of optics, mechanics and astronomy began to be applied to questions raised by natural philosophy.<sup>[60]</sup> Grant argues that Newton attempted to expose the mathematical basis of nature – the immutable rules it obeyed – and in doing so joined natural philosophy and mathematics for the first time, producing an early work of modern physics.<sup>[61]</sup>

The scientific revolution, which began to take hold in the 1600s, represented a sharp break from Aristotelian modes of inquiry.<sup>[62]</sup> One of its principal advances was the use of the scientific method to investigate nature. Data was collected and repeatable measurements made in experiments.<sup>[63]</sup> Scientists then formed hypotheses to explain the results of these experiments.<sup>[64]</sup> The hypothesis was then tested using the principle of falsifiability to prove or disprove its accuracy.<sup>[64]</sup> The natural sciences continued to be called natural philosophy, but the adoption of the scientific method took science beyond the realm of philosophical conjecture and introduced a more structured way of examining nature.<sup>[62]</sup>

Newton, an English mathematician and physicist, was the seminal figure in the scientific revolution.<sup>[65]</sup> Drawing on advances made in astronomy by Copernicus, Brahe and Kepler, Newton derived the universal law of gravitation and laws of motion.<sup>[66]</sup> These laws applied both on earth and in outer space, uniting two spheres of the physical world previously thought to function independently of each other, according to separate physical rules.<sup>[67]</sup> Newton, for example, showed that the tides were caused by the gravitational pull of the moon.<sup>[68]</sup> Another of Newton's advances was to make mathematics a powerful explanatory tool for natural phenomena.<sup>[69]</sup> While natural philosophers had long used mathematics as a means of measurement and analysis, its principles were not used as a means of understanding cause and effect in nature until Newton.<sup>[69]</sup>

In the 1700s and 1800s, scientists including Charles-Augustin de Coulomb, Alessandro Volta, and Michael Faraday built upon Newtonian mechanics by exploring electromagnetism, or the interplay of forces with positive and negative charges on electrically charged particles.<sup>[70]</sup> Faraday proposed that forces in nature operated in "fields" that filled space.<sup>[71]</sup> The idea of fields contrasted with the Newtonian construct of gravitation as simply "action at a distance", or the attraction of objects with nothing in the space between them to intervene.<sup>[71]</sup> James Clerk Maxwell in the 19th century unified these discoveries in a coherent theory of electrodynamics.<sup>[70]</sup> Using mathematical equations and experimentation, Maxwell discovered that space was filled with charged particles that could act upon themselves and each other, and that they were a medium for the transmission of charged waves.<sup>[70]</sup>

Significant advances in chemistry also took place during the scientific revolution. Antoine Lavoisier, a French chemist, refuted the phlogiston theory, which posited that things burned by releasing "phlogiston" into the air.<sup>[71]</sup> Joseph Priestley had discovered oxygen in the 1700s, but Lavoisier discovered that combustion was the result of oxidation.<sup>[71]</sup> He also constructed a table of 33 elements and invented modern chemical nomenclature.<sup>[71]</sup> Formal biological science remained in its infancy in the 18th century, when the focus lay upon the classification and categorization of natural life. This growth in natural history was led by Carolus Linnaeus, whose 1735 taxonomy of the natural world is still in use. Linnaeus in the 1750s introduced scientific names for all his species.<sup>[72]</sup>

## 19th-century developments (1800–1900)

By the 19th century, the study of science had come into the purview of professionals and institutions. In so doing, it gradually acquired the more modern name of *natural science*. The term *scientist* was coined by William Whewell in an 1834 review of Mary Somerville's *On the Connexion of the Sciences*.<sup>[73]</sup> But the word did not enter general use until nearly the end of the same century.

## Modern natural science (1900–present)

According to a famous 1923 textbook *Thermodynamics and the Free Energy of Chemical Substances* by the American chemist Gilbert N. Lewis and the American physical chemist Merle Randall,<sup>[74]</sup> the natural sciences contain three great branches:

Aside from the logical and mathematical sciences, there are three great branches of *natural science* which stand apart by reason of the variety of far reaching deductions drawn from a small number of primary postulates — they are mechanics, electrodynamics, and thermodynamics.<sup>[75]</sup>

Today, natural sciences are more commonly divided into life sciences, such as botany and zoology; and physical sciences, which include physics, chemistry, geology and astronomy.

## Branches of natural science

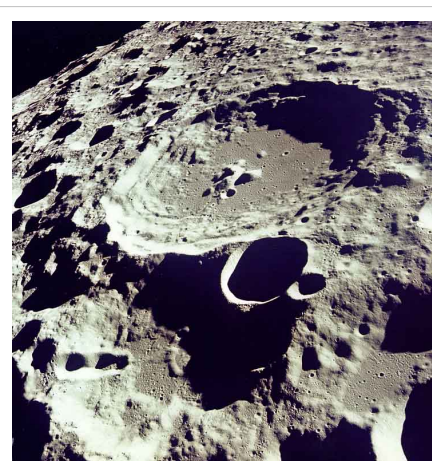
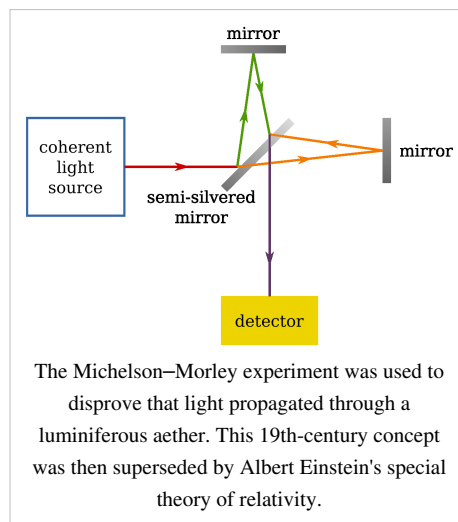
### Astronomy

This discipline is the science of celestial objects and phenomena that originate outside the Earth's atmosphere. It is concerned with the evolution, physics, chemistry, meteorology, and motion of celestial objects, as well as the formation and development of the universe.

Astronomy includes the examination, study and modeling of stars, planets, comets, galaxies and the cosmos. Most of the information used by astronomers is gathered by remote observation, although some laboratory reproduction of celestial phenomenon has been performed (such as the molecular chemistry of the interstellar medium).

While the origins of the study of celestial features and phenomenon can be traced back to antiquity, the scientific methodology of this field began to develop in the middle of the 17th century. A key factor was Galileo's introduction of the telescope to examine the night sky in more detail.

The mathematical treatment of astronomy began with Newton's development of celestial mechanics and the laws of gravitation, although it was triggered by earlier work of astronomers such as Kepler. By the 19th century, astronomy had developed into a formal science, with the introduction of instruments such as the spectroscope and photography, along with much-improved telescopes and the creation of professional observatories.



Space missions have been used to image distant locations within the Solar System, such as this *Apollo 11* view of Daedalus crater on the far side of the Moon.

## Biology

This field encompasses a set of disciplines that examines phenomena related to living organisms. The scale of study can range from sub-component biophysics up to complex ecologies. Biology is concerned with the characteristics, classification and behaviors of organisms, as well as how species were formed and their interactions with each other and the environment.

The biological fields of botany, zoology, and medicine date back to early periods of civilization, while microbiology was introduced in the 17th century with the invention of the microscope. However, it was not until the 19th century that biology became a unified science. Once scientists discovered commonalities between all living things, it was decided they were best studied as a whole.

Some key developments in biology were the discovery of genetics; Darwin's theory of evolution through natural selection; the germ theory of disease and the application of the techniques of chemistry and physics at the level of the cell or organic molecule.

Modern biology is divided into subdisciplines by the type of organism and by the scale being studied. Molecular biology is the study of the fundamental chemistry of life, while cellular biology is the examination of the cell; the basic building block of all life. At a higher level, physiology looks at the internal structure of organism, while ecology looks at how various organisms interrelate.

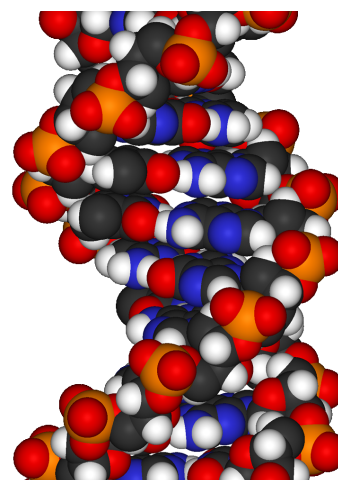
## Chemistry

Constituting the scientific study of matter at the atomic and molecular scale, chemistry deals primarily with collections of atoms, such as gases, molecules, crystals, and metals. The composition, statistical properties, transformations and reactions of these materials are studied. Chemistry also involves understanding the properties and interactions of individual atoms for use in larger-scale applications.

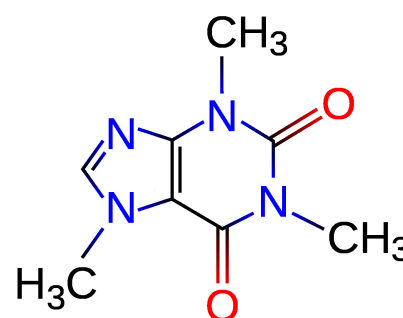
Most chemical processes can be studied directly in a laboratory, using a series of (often well-tested) techniques for manipulating materials, as well as an understanding of the underlying processes. Chemistry is often called "the central science" because of its role in connecting the other natural sciences.

Early experiments in chemistry had their roots in the system of Alchemy, a set of beliefs combining mysticism with physical experiments. The science of chemistry began to develop with the work of Robert Boyle, the discoverer of gas, and Antoine Lavoisier, who developed the theory of the Conservation of mass.

The discovery of the chemical elements and the concept of Atomic Theory began to systematize this science, and researchers developed a fundamental understanding of states of matter, ions, chemical bonds and chemical reactions. The success of this science led to a complementary chemical industry that now plays a significant role in the world economy.



A fragment of DNA, the chemical sequence that contains genetic instructions for the development and functioning of living organisms



This structural formula for molecule caffeine shows a graphical representation of how the atoms are arranged.

## Materials science

Originally developed through the field of metallurgy, the study of the properties of materials has now expanded into many materials other than metals. The field covers the chemistry, physics and engineering applications of materials including metals, ceramics, artificial polymers, and many others.

## Physics

Physics embodies the study of the fundamental constituents of the universe, the forces and interactions they exert on one another, and the results produced by these interactions. In general, physics is regarded as the fundamental science, because all other natural sciences use and obey the principles and laws set down by the field. Physics relies heavily on mathematics as the logical framework for formulation and quantification of principles.

The study of the principles of the universe has a long history and largely derives from direct observation and experimentation. The formulation of theories about the governing laws of the universe has been central to the study of physics from very early on, with philosophy gradually yielding to systematic, quantitative experimental testing and observation as the source of verification. Key historical developments in physics include Isaac Newton's theory of universal gravitation and classical mechanics, an understanding of electricity and its relation to magnetism, Einstein's theories of special and general relativity, the development of thermodynamics, and the quantum mechanical model of atomic and subatomic physics.

The field of physics is extremely broad, and can include such diverse studies as quantum mechanics and theoretical physics, applied physics and optics. Modern physics is becoming increasingly specialized, where researchers tend to focus on a particular area rather than being "universalists" like Isaac Newton, Albert Einstein and Lev Landau, who worked in multiple areas.

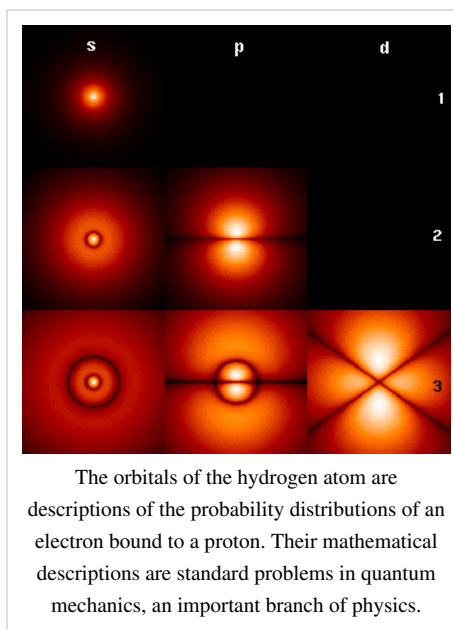
## Earth science

Earth science (also known as geoscience), is an all-embracing term for the sciences related to the planet Earth, including geology, geophysics, hydrology, meteorology, physical geography, oceanography, and soil science.

Although mining and precious stones have been human interests throughout the history of civilization, the development of the related sciences of economic geology and mineralogy did not occur until the 18th century. The study of the earth, particularly palaeontology, blossomed in the 19th century. The growth of other disciplines, such as geophysics, in the 20th century led to the development of the theory of plate tectonics in the 1960s, which has had a similar effect on the Earth sciences as the theory of evolution had on biology. Earth sciences today are closely linked to petroleum and mineral resources, climate research and to environmental assessment and remediation.

## Atmospheric science

Though sometimes considered in conjunction with the earth sciences, due to the independent development of its concepts, techniques and practices and also the fact of it having a wide range of sub disciplines under its wing, the atmospheric science is also considered a separate branch of natural science. This field studies the characteristics of different layers of the atmosphere from ground level to the edge of the time. The timescale of study also varies from days to centuries. Sometimes the field also includes the study of climatic patterns on planets other than earth.



## Oceanography

The serious study of oceans began in the early to mid-1900s. As a field of natural science, it is relatively young but stand-alone programs offer specializations in the subject. Though some controversies remain as to the categorization of the field under earth sciences, interdisciplinary sciences or as a separate field in its own right, most modern workers in the field agree that it has matured to a state that it has its own paradigms and practices. As such a big family of related studies spanning every aspect of the oceans is now classified under this field.

## Interdisciplinary studies

The distinctions between the natural science disciplines are not always sharp, and they share a number of cross-discipline fields. Physics plays a significant role in the other natural sciences, as represented by astrophysics, geophysics, chemical physics and biophysics. Likewise chemistry is represented by such fields as biochemistry, geochemistry and astrochemistry.

A particular example of a scientific discipline that draws upon multiple natural sciences is environmental science. This field studies the interactions of physical, chemical, geological, and biological components of the environment, with a particular regard to the effect of human activities and the impact on biodiversity and sustainability. This science also draws upon expertise from other fields such as economics, law and social sciences.

A comparable discipline is oceanography, as it draws upon a similar breadth of scientific disciplines. Oceanography is sub-categorized into more specialized cross-disciplines, such as physical oceanography and marine biology. As the marine ecosystem is very large and diverse, marine biology is further divided into many subfields, including specializations in particular species.

There are also a subset of cross-disciplinary fields which, by the nature of the problems that they address, have strong currents that run counter to specialization. Put another way: In some fields of integrative application, specialists in more than one field are a key part of most dialog. Such integrative fields, for example, include nanoscience, astrobiology, and complex system informatics.

## References

- [1] Ledoux 2002, p. 34
- [2] Barr 2006, p. 1
- [3] Simhony 1994, p. 49
- [4] Oglivie 2008, pp. 1–2
- [5] "Natural History" (<http://wordnetweb.princeton.edu/perl/webwn?s=natural+history>). Princeton University WordNet. . Retrieved October 21, 2012.
- [6] Lagemaat 2006, p. 283
- [7] Grant 2007, p. 1
- [8] Grant 2007, p. 2
- [9] Grant 2007, pp. 2–3
- [10] Magner 2002, p. 3
- [11] Magner 2002, pp. 3–4
- [12] Magner 2002, p. 4
- [13] Magner 2002, p. 5
- [14] Grant 2007, p. 8
- [15] Barr 2006, p. 2
- [16] Barr 2006, p. 3
- [17] Grant 2007, pp. 21–22
- [18] Grant 2007, pp. 27–28
- [19] Grant 2007, pp. 33–34
- [20] Grant 2007, p. 34
- [21] Grant 2007, pp. 34–35
- [22] Grant 2007, pp. 37–39, 53
- [23] Grant 2007, p. 52

- [24] Grant 2007, p. 95
- [25] Grant 2007, pp. 54, 59
- [26] Grant 2007, p. 103
- [27] Grant 2007, pp. 61–66
- [28] Barr 2006, p. 11
- [29] Barr 2006, pp. 11–12
- [30] Grant 2007, pp. 95, 130
- [31] Grant 2007, p. 106
- [32] Grant 2007, pp. 106–107
- [33] Grant 2007, p. 115
- [34] Grant 2007, p. 130
- [35] Grant 2007, p. 143
- [36] Grant 2007, p. 155
- [37] Grant 2007, p. 156
- [38] Grant 2007, pp. 156–157
- [39] Grant 2007, p. 158
- [40] Grant 2007, pp. 159–163
- [41] Grant 2007, p. 234
- [42] Grant 2007, pp. 236–237
- [43] Grant 2007, pp. 170–178
- [44] Grant 2007, pp. 189–190
- [45] Grant 2007, pp. 239–240
- [46] Grant 2007, pp. 241–243
- [47] Grant 2007, pp. 246–247
- [48] Grant 2007, p. 251
- [49] Grant 2007, p. 252
- [50] Grant 2007, p. 274
- [51] Grant 2007, p. 274–275
- [52] Grant 2007, pp. 276–277
- [53] Grant 2007, p. 278
- [54] Grant 2007, pp. 278–279
- [55] Grant 2007, p. 279
- [56] Grant 2007, pp. 280–285
- [57] Grant 2007, pp. 280–290
- [58] Grant 2007, pp. 280–295
- [59] Grant 2007, pp. 304–306
- [60] Grant 2007, p. 307
- [61] Grant 2007, pp. 317–318
- [62] Barr 2006, p. 26
- [63] Barr 2006, pp. 26–27
- [64] Barr 2006, p. 27
- [65] Barr 2006, p. 33
- [66] Barr 2006, pp. 33–35
- [67] Barr 2006, p. 35
- [68] Barr 2006, p. 36
- [69] Barr 2006, p. 37
- [70] Barr 2006, p. 48
- [71] Barr 2006, p. 49
- [72] Mayr 1982, pp. 171–179
- [73] Holmes, R (2008). *The age of wonder: How the romantic generation discovered the beauty and terror of science*. London: Harper Press. p. 449. ISBN 9780007149537.
- [74] Lewis, Gilbert N.; Randall, Merle (1923). *Thermodynamics and the Free Energy of Chemical Substances*. later Printing edition (First ed.). McGraw-Hill Book Company. ASIN B000GSLHZS.
- [75] Huggins, Robert A. (2010). *Energy storage* (Online-Ausg. ed.). New York: Springer. p. 13. ISBN 9781441910233.
-

## Bibliography

- Barr, Stephen M. (2006). *A Students Guide to Natural Science*. Wilmington, DE: Intercollegiate Studies Institute. ISBN 978-1-932-23692-7.
- Gohau, Gabriel (1990). *A History of Geology*. Revised and translated by Albert V. Carozzi and Marguerite Carozzi. New Brunswick: Rutgers University Press. ISBN 978-0-8135-1666-0.
- Grant, Edward (2007). *A History of Natural Philosophy: From the Ancient World to the Nineteenth Century*. Cambridge: Cambridge University Press. ISBN 978-0-521-68957-1.
- Lagemaat, Ricahrd van de (2006). *Theory of Knowledge for the IB Diploma* ([http://books.google.ae/books?id=l\\_HMz6Ub-rcC&pg=PA283&lpg=PA283&dq=hard+science+physics+chemistry&source=bl&ots=MkgFAdzFsK&sig=SxaUdq4ltk-yiXS\\_FmrdbsBLICU&hl=en&sa=X&ei=cgmEUKfKAc3yrQfuu4GIDg&ved=0CDwQ6AEwAw#v=onepage&q=hard+science+physics+chemistry&f=false](http://books.google.ae/books?id=l_HMz6Ub-rcC&pg=PA283&lpg=PA283&dq=hard+science+physics+chemistry&source=bl&ots=MkgFAdzFsK&sig=SxaUdq4ltk-yiXS_FmrdbsBLICU&hl=en&sa=X&ei=cgmEUKfKAc3yrQfuu4GIDg&ved=0CDwQ6AEwAw#v=onepage&q=hard+science+physics+chemistry&f=false)). Cambridge: Cambridge University Press. ISBN 978-0-521-54298-2.
- Ledoux, Stephen F. (2002). "Defining Natural Sciences" (<http://behaviorology.org/pdf/DefineNatlSciences.pdf>) (PDF). *Behaviorology Today* **5** (1): 34. "Fundamentally, natural sciences are defined as disciplines that deal only with natural events (i.e., independent and dependent variables in nature) using scientific methods."
- Magner, Lois N. (2002). *A History of the Life Sciences*. New York: Marcel Dekker, Inc.. ISBN 978-0-8247-0824-5.
- Mayr, Ernst (1982). *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press. ISBN 978-0-674-36445-7.
- Oglivie, Brian W. (2008). *The Science of Describing: Natural History in Renaissance Europe*. Chicago: University of Chicago Press. ISBN 978-0-226-62088-6.
- Prpic, Katarina (2009). *Beyond the Myths about the Natural and Social Sciences: A Sociological View*. Zagreb: Institute for Social Research. ISBN 978-953-6218-40-0.
- Simhony, M. (2006). *Invitation to the Natural Physics of Matter, Space, and Radiation* ([http://books.google.ae/books?id=C6tEf\\_P2rdYC&pg=PA49&lpg=PA49&dq="natural+sciences"+biology+astronomy+physics+chemistry&source=bl&ots=4Yshn5Dbwn&sig=Xz0m6-d8IkNw9C-edfUZXsmcGL8&hl=en&sa=X&ei=ityCUKCwLNDPrQfExoGgBQ&ved=0CCkQ6AEwAA#v=onepage&q="natural+sciences"+biology+astronomy+physics+chemistry&f=false](http://books.google.ae/books?id=C6tEf_P2rdYC&pg=PA49&lpg=PA49&dq=)). Singapore: World Scientific Publishing Co., Inc.. ISBN 978-9-810-21649-8.
- Smith, C.H. Llewellyn (1997). "The use of basic science" (<http://public.web.cern.ch/public/en/about/BasicScience1-en.html>). CERN. Retrieved October 20, 2012.
- Stokes, Donald E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation* ([http://books.google.ae/books?id=xbzfTkGKOHEC&printsec=frontcover&dq=fundamental+science+and+applied+science&source=bl&ots=qhRzhXdF\\_n&sig=QblARvDVIX0XMqW3N9G-L4nhINE&hl=en&sa=X&ei=g\\_iCUPfQAZCuraFzwoGYDg&ved=0CCkQ6AEwAA#v=onepage&q=fundamental+science+and+applied+science&f=false](http://books.google.ae/books?id=xbzfTkGKOHEC&printsec=frontcover&dq=fundamental+science+and+applied+science&source=bl&ots=qhRzhXdF_n&sig=QblARvDVIX0XMqW3N9G-L4nhINE&hl=en&sa=X&ei=g_iCUPfQAZCuraFzwoGYDg&ved=0CCkQ6AEwAA#v=onepage&q=fundamental+science+and+applied+science&f=false)). Washington, D.C.: Brookings Institution Press. ISBN 978-0-815-78177-6.



## Further reading

- Defining Natural Sciences (<http://www.behaviorology.org/pdf/DefineNatlSciences.pdf>) Ledoux, S. F., 2002: Defining Natural Sciences, *Behaviorology Today*, **5(1)**, 34-36.

## External links

- The History of Recent Science and Technology (<http://hrst.mit.edu/>)
- Natural Sciences (<http://www.dur.ac.uk/natural.sciences/>) Information on the Natural Sciences degree programme at Durham University.
- Natural Sciences (<http://sciencia.com/>) Contains updated information on research in the Natural Sciences including biology, geography and the applied life and earth sciences.
- Natural Sciences (<http://www.bath.ac.uk/nat-sci/>) Information on the Natural Sciences degree programme at the University of Bath which includes the Biological Sciences, Chemistry, Pharmacology, Physics and Environmental Studies.
- Reviews of Books About Natural Science (<http://www.scibooks.org/>) This site contains over 50 previously published reviews of books about natural science, plus selected essays on timely topics in natural science.
- Scientific Grant Awards Database (<http://search.engrant.com/>) Contains details of over 2,000,000 scientific research projects conducted over the past 25 years.
- Natural Sciences Tripos (<http://www.cam.ac.uk/about/natscitripos/>) Provides information on the framework within which most of the natural science is taught at the University of Cambridge.
- E!Science (<http://esciencenews.com/sources>) Up-to-date science news aggregator from major sources including universities.

## Scientific method

---

**The scientific method** (or simply **scientific method**) is a body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge.<sup>[1]</sup> To be termed scientific, a method of inquiry must be based on empirical and measurable evidence subject to specific principles of reasoning.<sup>[2]</sup> The *Oxford English Dictionary* defines the scientific method as: "a method or procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses."<sup>[3]</sup>

The chief characteristic which distinguishes the scientific method from other methods of acquiring knowledge is that scientists seek to let reality speak for itself, supporting a theory when a theory's predictions are confirmed and challenging a theory when its predictions prove false. Although procedures vary from one field of inquiry to another, identifiable features distinguish scientific inquiry from other methods of obtaining knowledge. Scientific researchers propose hypotheses as explanations of phenomena, and design experimental studies to test these hypotheses via predictions which can be derived from them. These steps must be repeatable, to guard against mistake or confusion in any particular experimenter. Theories that encompass wider domains of inquiry may bind many independently derived hypotheses together in a coherent, supportive structure. Theories, in turn, may help form new hypotheses or place groups of hypotheses into context.

Scientific inquiry is generally intended to be as objective as possible in order to reduce biased interpretations of results. Another basic expectation is to document, archive and share all data and methodology so they are available for careful scrutiny by other scientists, giving them the opportunity to verify results by attempting to reproduce them. This practice, called *full disclosure*, also allows statistical measures of the reliability of these data to be established (when data is sampled or compared to chance).

---

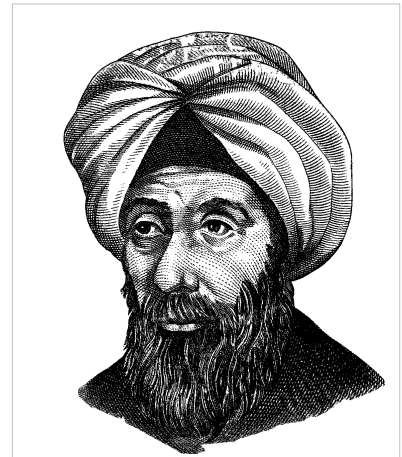
## Overview

Scientific method has been practiced in some form for at least one thousand years<sup>[4]</sup> and is the process by which science is carried out. Because science builds on previous knowledge, it consistently improves our understanding of the world. The scientific method also improves itself in the same way, meaning that it gradually becomes more effective at generating new knowledge. For example, the concept of falsification (first proposed in 1934) reduces confirmation bias by formalizing the attempt to *disprove* hypotheses rather than prove them.<sup>[8]</sup>

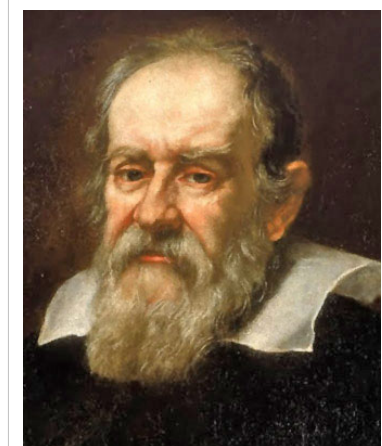
The overall process involves making conjectures (hypotheses), deriving predictions from them as logical consequences, and then carrying out experiments based on those predictions to determine whether the original conjecture was correct. There are difficulties in a formulaic statement of method, however. Though the scientific method is often presented as a fixed sequence of steps, they are better considered as general principles.<sup>[9]</sup> Not all steps take place in every scientific inquiry (or to the same degree), and not always in the same order. As noted by William Whewell (1794–1866), "invention, sagacity, [and] genius"<sup>[10]</sup> are required at every step:

**Formulate a question:** The question can refer to the explanation of a specific *observation*, as in "Why is the sky blue?", but can also be open-ended, as in "Does sound travel faster in air than in water?" or "How can I design a drug to cure this particular disease?" This stage also involves looking up and evaluating previous evidence from other scientists, as well as considering one's own experience. If the answer is already known, a different question that builds on the previous evidence can be posed. When applying the scientific method to scientific research, determining a good question can be very difficult and affects the final outcome of the investigation.<sup>[11]</sup>

**Hypothesis:** An hypothesis is a conjecture, based on the knowledge obtained while formulating the question, that may explain the observed behavior of a part of our universe. The hypothesis might be very specific, e.g., Einstein's equivalence principle or Francis Crick's "DNA makes RNA makes protein",<sup>[12]</sup> or it might be broad, e.g., unknown species of life dwell in the unexplored depths of the oceans. A statistical hypothesis is a conjecture about some population. For example, the population might be people with a particular disease. The conjecture might be that a new drug will cure the disease in some of those people. Terms commonly associated with statistical hypotheses are null hypothesis and alternative hypothesis. A null hypothesis is the conjecture that the statistical hypothesis is false, e.g., that the new drug does nothing and that any cures are due to chance effects. Researchers normally want to show that the null hypothesis is false. The alternative hypothesis is the desired outcome, e.g., that the drug does better than



Ibn al-Haytham (Alhazen), 965–1039 Iraq. The Arab scholar who lived during the Islamic golden age is considered by some to be the father of modern scientific methodology.<sup>[4]</sup>



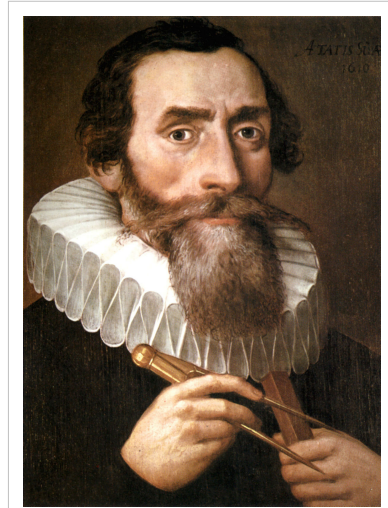
According to Morris Kline,<sup>[5]</sup> "Modern science owes its present flourishing state to a new scientific method which was fashioned almost entirely by Galileo Galilei." Dudley Shapere<sup>[6]</sup> takes a more measured view of Galileo's contribution.

chance. A final point: a scientific hypothesis must be falsifiable, meaning that one can identify a possible outcome of an experiment that conflicts with predictions deduced from the hypothesis; otherwise, it cannot be meaningfully tested.

**Prediction:** This step involves determining the logical consequences of the hypothesis. One or more predictions are then selected for further testing. The less likely that the prediction would be correct simply by coincidence, the stronger evidence it would be if the prediction were fulfilled; evidence is also stronger if the answer to the prediction is not already known, due to the effects of hindsight bias (see also postdiction). Ideally, the prediction must also distinguish the hypothesis from likely alternatives; if two hypotheses make the same prediction, observing the prediction to be correct is not evidence for either one over the other. (These statements about the relative strength of evidence can be mathematically derived using Bayes' Theorem.)

**Test:** This is an investigation of whether the real world behaves as predicted by the hypothesis. Scientists (and other people) test hypotheses by conducting experiments. The purpose of an experiment is to determine whether observations of the real world agree with or conflict with the predictions derived from an hypothesis. If they agree, confidence in the hypothesis increases; otherwise, it decreases. Agreement does not assure that the hypothesis is true; future experiments may reveal problems. Karl Popper advised scientists to try to falsify hypotheses, i.e., to search for and test those experiments that seem most doubtful. Large numbers of successful confirmations are not convincing if they arise from experiments that avoid risk.<sup>[13]</sup> Experiments should be designed to minimize possible errors, especially through the use of appropriate scientific controls. For example, tests of medical treatments are commonly run as double-blind tests. Test personnel, who might unwittingly reveal to test subjects which samples are the desired test drugs and which are placebos, are kept ignorant of which are which. Such hints can bias the responses of the test subjects. Failure of an experiment does not necessarily mean the hypothesis is false. Experiments always depend on several hypotheses, e.g., that the test equipment is working properly, and a failure may be a failure of one of the auxiliary hypotheses. (See the Duhem-Quine thesis.) Experiments can be conducted in a college lab, on a kitchen table, at CERN's Large Hadron Collider, at the bottom of an ocean, on Mars (using one of the working rovers), and so on. Astronomers do experiments, searching for planets around distant stars. Finally, most individual experiments address highly specific topics for reasons of practicality. As a result, evidence about broader topics is usually accumulated gradually.

**Analysis:** This involves determining what the results of the experiment show and deciding on the next actions to take. The predictions of the hypothesis are compared to those of the null hypothesis, to determine which is better able to explain the data. In cases where an experiment is repeated many times, a statistical analysis such as a chi-squared test may be required. If the evidence has falsified the hypothesis, a new hypothesis is required; if the experiment supports the hypothesis but the evidence is not strong enough for high confidence, other predictions from the hypothesis must be tested. Once a hypothesis is strongly supported by evidence, a new question can be asked to provide further insight on the same topic. Evidence from other scientists and one's own experience can be incorporated at any stage in the process. Many iterations may be required to gather sufficient evidence to answer a question with confidence, or to build up many answers to highly specific questions in order to answer a single broader question.



Johannes Kepler (1571–1630). "Kepler shows his keen logical sense in detailing the whole process by which he finally arrived at the true orbit. This is the greatest piece of Retroductive reasoning ever performed." —C. S. Peirce, c. 1896, on Kepler's reasoning through explanatory hypotheses<sup>[7]</sup>

This model underlies the scientific revolution. One thousand years ago, Alhazen demonstrated the importance of forming questions and subsequently testing them,<sup>[4]</sup> an approach which was advocated by Galileo in 1638 with the publication of *Two New Sciences*.<sup>[14]</sup> The current method is based on a hypothetico-deductive model<sup>[15]</sup> formulated in the 20th century, although it has undergone significant revision since first proposed (for a more formal discussion, see below).

## DNA example



The basic elements of the scientific method are illustrated by the following example from the discovery of the structure of DNA:

- **Question:** Previous investigation of DNA had determined its chemical composition (the four nucleotides), the structure of each individual nucleotide, and other properties. It had been identified as the carrier of genetic information by the Avery–MacLeod–McCarty experiment in 1944,<sup>[16]</sup> but the mechanism of how genetic information was stored in DNA was unclear.
- **Hypothesis:** Francis Crick and James D. Watson hypothesized that DNA had a helical structure.<sup>[17]</sup>
- **Prediction:** If DNA had a helical structure, its X-ray diffraction pattern would be X-shaped.<sup>[18][19]</sup> This prediction was determined using the mathematics of the helix transform, which had been derived by Cochran, Crick and Vand<sup>[20]</sup> (and independently by Stokes).
- **Experiment:** Rosalind Franklin crystallized pure DNA and performed X-ray diffraction to produce photo 51. The results showed an X-shape.
- **Analysis:** When Watson saw the detailed diffraction pattern, he immediately recognized it as a helix.<sup>[21][22]</sup> He and Crick then produced their model, using this information along with the previously known information about DNA's composition and about molecular interactions such as hydrogen bonds.<sup>[23]</sup>

The discovery became the starting point for many further studies involving the genetic material, such as the field of molecular genetics, and it was awarded the Nobel Prize in 1962. Each step of the example is examined in more detail later in the article.

## Other components

The scientific method also includes other components required even when all the iterations of the steps above have been completed:

**Replication:** If an experiment cannot be repeated to produce the same results, this implies that the original results were in error. As a result, it is common for a single experiment to be performed multiple times, especially when there are uncontrolled variables or other indications of experimental error. For significant or surprising results, other scientists may also attempt to replicate the results for themselves, especially if those results would be important to their own work.

**External review:** The process of peer review involves evaluation of the experiment by experts, who give their opinions anonymously to allow them to give unbiased criticism. It does not certify correctness of the results, only that the experiments themselves were sound (based on the description supplied by the experimenter). If the work passes peer review, which may require new experiments requested by the reviewers, it will be published in a peer-reviewed scientific journal. The specific journal that publishes the results indicates the perceived quality of the work.

**Data recording and sharing:** Scientists must record all data very precisely in order to reduce their own bias and aid in replication by others, a requirement first promoted by Ludwik Fleck (1896–1961) and others.<sup>[24]</sup> They must supply this data to other scientists who wish to replicate any results, extending to the sharing of any experimental samples that may be difficult to obtain.<sup>[25]</sup>

## Scientific inquiry

The goal of a scientific inquiry is to obtain knowledge in the form of testable explanations that can predict the results of future experiments. This allows scientists to gain an understanding of reality, and later use that understanding to intervene in its causal mechanisms (such as to cure disease). The better an explanation is at making predictions, the more useful it is, and the more likely it is to be correct. The most successful explanations, which explain and make accurate predictions in a wide range of circumstances, are called scientific theories.

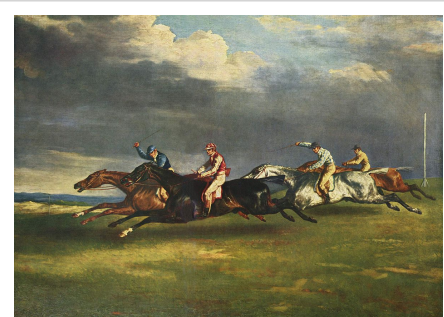
Most experimental results do not result in large changes in human understanding; improvements in theoretical scientific understanding is usually the result of a gradual synthesis of the results of different experiments, by various researchers, across different domains of science.<sup>[26]</sup> Scientific models vary in the extent to which they have been experimentally tested and for how long, and in their acceptance in the scientific community. In general, explanations become accepted by a scientific community as evidence in favor is presented, and as presumptions that are inconsistent with the evidence are falsified.

## Properties of scientific inquiry

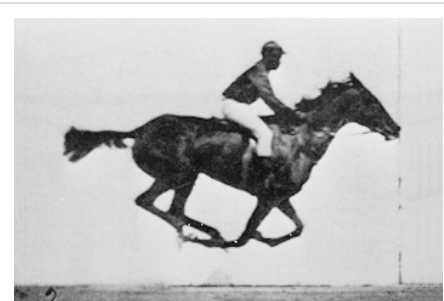
Scientific knowledge is closely tied to empirical findings, and always remains subject to falsification if new experimental observation incompatible with it is found. That is, no theory can ever be considered completely certain, since new evidence falsifying it might be discovered. If such evidence is found, a new theory may be proposed, or (more commonly) it is found that minor modifications to the previous theory are sufficient to explain the new evidence. The strength of a theory is related to how long it has persisted without falsification of its core principles.

Confirmed theories are also subject to subsumption by more accurate theories. For example, thousands of years of scientific observations of the planets were explained almost perfectly by Newton's laws. However, these laws were then determined to be special cases of a more general theory (relativity), which explained both the (previously unexplained) exceptions to Newton's laws as well as predicting and explaining other observations such as the deflection of light by gravity. Thus independent, unconnected, scientific observations can be connected to each other, unified by principles of increasing explanatory power.<sup>[27]</sup>

Since every new theory must explain even more than the previous one, any successor theory capable of subsuming it must meet an even higher standard, explaining both the larger, unified body of observations explained by the previous theory and unifying that with even more observations. In other words, as scientific knowledge becomes more accurate with time, it becomes increasingly harder to produce a more successful theory, simply because of the great success of the theories that already exist.<sup>[27]</sup> For example, the Theory of Evolution explains the diversity of life on Earth, how species adapt to their environments, and many other patterns observed in the natural world,<sup>[28][29]</sup> its most recent major modification was unification with genetics to form the modern evolutionary synthesis. In subsequent modifications, it has also subsumed aspects of many other fields such as biochemistry and molecular biology.



Flying gallop falsified; see image below.



Eadweard Muybridge's studies of a horse galloping

## Beliefs and biases

Scientific methodology directs that hypotheses be tested in controlled conditions which can be reproduced by others. The scientific community's pursuit of experimental control and reproducibility diminishes the effects of cognitive biases.

For example, pre-existing beliefs can alter the interpretation of results, as in confirmation bias; this is a heuristic that leads a person with a particular belief to see things as reinforcing their belief, even if another observer might disagree (in other words, people tend to observe what they expect to observe).

A historical example is the conjecture that the legs of a galloping horse are splayed at the point when none of the horse's legs touches the ground, to the point of this image being included in paintings by its supporters. However, the first stop-action pictures of a horse's gallop by Eadweard Muybridge showed this to be false, and that the legs are instead gathered together.<sup>[30]</sup>

Another important human bias that plays a role is a preference for new, surprising statements (see appeal to novelty), which can result in a search for evidence that the new is true.<sup>[1]</sup>

In contrast to the requirement for scientific knowledge to correspond to reality, beliefs based on myth or stories can be believed and acted upon irrespective of truth,<sup>[31]</sup> often taking advantage of the narrative fallacy that when narrative is constructed its elements become easier to believe.<sup>[32][33]</sup> Myths intended to be taken as true must have their elements assumed *a priori*, while science requires testing and validation *a posteriori* before ideas are accepted.

## Elements of the scientific method

There are different ways of outlining the basic method used for scientific inquiry. The scientific community and philosophers of science generally agree on the following classification of method components. These methodological elements and organization of procedures tend to be more characteristic of natural sciences than social sciences. Nonetheless, the cycle of formulating hypotheses, testing and analyzing the results, and formulating new hypotheses, will resemble the cycle described below.

Four essential elements<sup>[34][35][36]</sup> of the scientific method<sup>[37]</sup> are iterations,<sup>[38][39]</sup> recursions,<sup>[40]</sup> interleavings, or orderings of the following:

- Characterizations (observations,<sup>[41]</sup> definitions, and measurements of the subject of inquiry)
- Hypotheses<sup>[42][43]</sup> (theoretical, hypothetical explanations of observations and measurements of the subject)<sup>[44]</sup>
- Predictions (reasoning including logical deduction<sup>[45]</sup> from the hypothesis or theory)
- Experiments<sup>[46]</sup> (tests of all of the above)

Each element of the scientific method is subject to peer review for possible mistakes. These activities do not describe all that scientists do (see below) but apply mostly to experimental sciences (e.g., physics, chemistry, and biology). The elements above are often taught in the educational system as "the scientific method".<sup>[47]</sup>

The scientific method is not a single recipe: it requires intelligence, imagination, and creativity.<sup>[48]</sup> In this sense, it is not a mindless set of standards and procedures to follow, but is rather an ongoing cycle, constantly developing more useful, accurate and comprehensive models and methods. For example, when Einstein developed the Special and General Theories of Relativity, he did not in any way refute or discount Newton's *Principia*. On the contrary, if the astronomically large, the vanishingly small, and the extremely fast are removed from Einstein's theories — all phenomena Newton could not have observed — Newton's equations are what remain. Einstein's theories are expansions and refinements of Newton's theories and, thus, increase our confidence in Newton's work.

A linearized, pragmatic scheme of the four points above is sometimes offered as a guideline for proceeding:<sup>[49]</sup>

1. Define a question
2. Gather information and resources (observe)
3. Form an explanatory hypothesis



4. Test the hypothesis by performing an experiment and collecting data in a reproducible manner
5. Analyze the data
6. Interpret the data and draw conclusions that serve as a starting point for new hypothesis
7. Publish results
8. Retest (frequently done by other scientists)

The iterative cycle inherent in this step-by-step method goes from point 3 to 6 back to 3 again.

While this schema outlines a typical hypothesis/testing method,<sup>[50]</sup> it should also be noted that a number of philosophers, historians and sociologists of science (perhaps most notably Paul Feyerabend) claim that such descriptions of scientific method have little relation to the ways science is actually practiced.

The "operational" paradigm combines the concepts of operational definition, instrumentalism, and utility:

The essential elements of scientific method are operations, observations, models, and a utility function for evaluating models.<sup>[51]</sup>

- Operation - Some action done to the system being investigated
- Observation - What happens when the operation is done to the system
- Model - A fact, hypothesis, theory, or the phenomenon itself at a certain moment
- Utility Function - A measure of the usefulness of the model to explain, predict, and control, and of the cost of use of it. One of the elements of any scientific utility function is the refutability of the model. Another is its simplicity, on the Principle of Parsimony more commonly known as Occam's Razor.

## Characterizations

The scientific method depends upon increasingly sophisticated characterizations of the subjects of investigation. (The *subjects* can also be called *unsolved problems* or the *unknowns*.) For example, Benjamin Franklin conjectured, correctly, that St. Elmo's fire was electrical in nature, but it has taken a long series of experiments and theoretical changes to establish this. While seeking the pertinent properties of the subjects, careful thought may also entail some definitions and observations; the observations often demand careful measurements and/or counting.

The systematic, careful collection of measurements or counts of relevant quantities is often the critical difference between pseudo-sciences, such as alchemy, and science, such as chemistry or biology. Scientific measurements are usually tabulated, graphed, or mapped, and statistical manipulations, such as correlation and regression, performed on them. The measurements might be made in a controlled setting, such as a laboratory, or made on more or less inaccessible or unmanipulatable objects such as stars or human populations. The measurements often require specialized scientific instruments such as thermometers, spectrosopes, particle accelerators, or voltmeters, and the progress of a scientific field is usually intimately tied to their invention and improvement.

"I am not accustomed to saying anything with certainty after only one or two observations."—Andreas Vesalius (1546)<sup>[52]</sup>

## Uncertainty

Measurements in scientific work are also usually accompanied by estimates of their uncertainty. The uncertainty is often estimated by making repeated measurements of the desired quantity. Uncertainties may also be calculated by consideration of the uncertainties of the individual underlying quantities used. Counts of things, such as the number of people in a nation at a particular time, may also have an uncertainty due to data collection limitations. Or counts may represent a sample of desired quantities, with an uncertainty that depends upon the sampling method used and the number of samples taken.



## Definition

Measurements demand the use of *operational definitions* of relevant quantities. That is, a scientific quantity is described or defined by how it is measured, as opposed to some more vague, inexact or "idealized" definition. For example, electrical current, measured in amperes, may be operationally defined in terms of the mass of silver deposited in a certain time on an electrode in an electrochemical device that is described in some detail. The operational definition of a thing often relies on comparisons with standards: the operational definition of "mass" ultimately relies on the use of an artifact, such as a particular kilogram of platinum-iridium kept in a laboratory in France.

The scientific definition of a term sometimes differs substantially from its natural language usage. For example, mass and weight overlap in meaning in common discourse, but have distinct meanings in mechanics. Scientific quantities are often characterized by their units of measure which can later be described in terms of conventional physical units when communicating the work.

New theories are sometimes developed after realizing certain terms have not previously been sufficiently clearly defined. For example, Albert Einstein's first paper on relativity begins by defining simultaneity and the means for determining length. These ideas were skipped over by Isaac Newton with, "*I do not define time, space, place and motion, as being well known to all.*" Einstein's paper then demonstrates that they (viz., absolute time and length independent of motion) were approximations. Francis Crick cautions us that when characterizing a subject, however, it can be premature to define something when it remains ill-understood.<sup>[53]</sup> In Crick's study of consciousness, he actually found it easier to study awareness in the visual system, rather than to study free will, for example. His cautionary example was the gene; the gene was much more poorly understood before Watson and Crick's pioneering discovery of the structure of DNA; it would have been counterproductive to spend much time on the definition of the gene, before them.

## DNA-characterizations

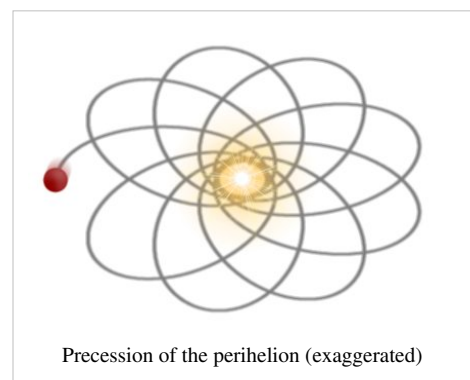


The history of the discovery of the structure of DNA is a classic example of the elements of the scientific method: in 1950 it was known that genetic inheritance had a mathematical description, starting with the studies of Gregor Mendel, and that DNA contained genetic information (Oswald Avery's *transforming principle*).<sup>[16]</sup> But the mechanism of storing genetic information (i.e., genes) in DNA was unclear.

Researchers in Bragg's laboratory at Cambridge University made X-ray diffraction pictures of various molecules, starting with crystals of salt, and proceeding to more complicated substances. Using clues painstakingly assembled over decades, beginning with its chemical composition, it was determined that it should be possible to characterize the physical structure of DNA, and the X-ray images would be the vehicle.<sup>[54]</sup> ..2. *DNA-hypotheses*

## Another example: precession of Mercury

The characterization element can require extended and extensive study, even centuries. It took thousands of years of measurements, from the Chaldean, Indian, Persian, Greek, Arabic and European astronomers, to fully record the motion of planet Earth. Newton was able to include those measurements into consequences of his laws of motion. But the perihelion of the planet Mercury's orbit exhibits a precession that cannot be fully explained by Newton's laws of motion (see diagram to the right), though it took quite some time to realize this. The observed difference for Mercury's precession between Newtonian theory and observation was one of the things that occurred to Einstein as a possible early test of his theory of General Relativity. His relativistic calculations matched observation much more closely than did Newtonian theory (the difference is approximately 43 arc-seconds per century), .



## Hypothesis development

An hypothesis is a suggested explanation of a phenomenon, or alternately a reasoned proposal suggesting a possible correlation between or among a set of phenomena.

Normally hypotheses have the form of a mathematical model. Sometimes, but not always, they can also be formulated as existential statements, stating that some particular instance of the phenomenon being studied has some characteristic and causal explanations, which have the general form of universal statements, stating that every instance of the phenomenon has a particular characteristic.

Scientists are free to use whatever resources they have — their own creativity, ideas from other fields, induction, Bayesian inference, and so on — to imagine possible explanations for a phenomenon under study. Charles Sanders Peirce, borrowing a page from Aristotle (*Prior Analytics*, 2.25) described the incipient stages of inquiry, instigated by the "irritation of doubt" to venture a plausible guess, as *abductive reasoning*. The history of science is filled with stories of scientists claiming a "flash of inspiration", or a hunch, which then motivated them to look for evidence to support or refute their idea. Michael Polanyi made such creativity the centerpiece of his discussion of methodology.

William Glen observes that

the success of a hypothesis, or its service to science, lies not simply in its perceived "truth", or power to displace, subsume or reduce a predecessor idea, but perhaps more in its ability to stimulate the research that will illuminate ... bald suppositions and areas of vagueness.<sup>[55]</sup>

In general scientists tend to look for theories that are "elegant" or "beautiful". In contrast to the usual English use of these terms, they here refer to a theory in accordance with the known facts, which is nevertheless relatively simple and easy to handle. Occam's Razor serves as a rule of thumb for choosing the most desirable amongst a group of equally explanatory hypotheses.

## DNA-hypotheses



Linus Pauling proposed that DNA might be a triple helix.<sup>[56]</sup> This hypothesis was also considered by Francis Crick and James D. Watson but discarded. When Watson and Crick learned of Pauling's hypothesis, they understood from existing data that Pauling was wrong<sup>[57]</sup> and that Pauling would soon admit his difficulties with that structure. So, the race was on to figure out the correct structure (except that Pauling did not realize at the time that he was in a race—see section on "DNA-predictions" below)

## Predictions from the hypothesis

Any useful hypothesis will enable predictions, by reasoning including deductive reasoning. It might predict the outcome of an experiment in a laboratory setting or the observation of a phenomenon in nature. The prediction can also be statistical and deal only with probabilities.

It is essential that the outcome of testing such a prediction be currently unknown. Only in this case does a successful outcome increase the probability that the hypothesis is true. If the outcome is already known, it's called a consequence and should have already been considered while formulating the hypothesis.

If the predictions are not accessible by observation or experience, the hypothesis is not yet testable and so will remain to that extent unscientific in a strict sense. A new technology or theory might make the necessary experiments feasible. Thus, much scientifically based speculation might convince one (or many) that the hypothesis that other intelligent species exist is true. But since there no experiment now known which can test this hypothesis, science itself can have little to say about the possibility. In future, some new technique might lead to an experimental test and the speculation would then become part of accepted science.

## DNA-predictions

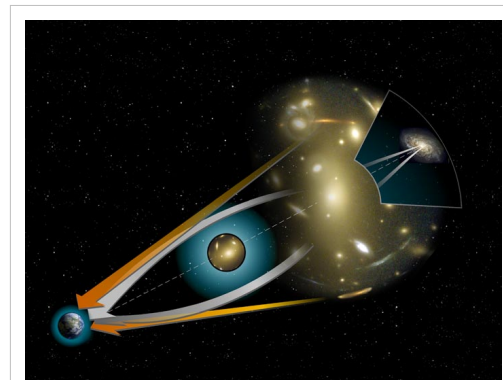


James D. Watson, Francis Crick, and others hypothesized that DNA had a helical structure. This implied that DNA's X-ray diffraction pattern would be 'x shaped'.<sup>[19][58]</sup> This prediction followed from the work of Cochran, Crick and Vand<sup>[20]</sup> (and independently by Stokes). The Cochran-Crick-Vand-Stokes theorem provided a mathematical explanation for the empirical observation that diffraction from helical structures produces x shaped patterns.

In their first paper, Watson and Crick also noted that the double helix structure they proposed provided a simple mechanism for DNA replication, writing "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material".<sup>[59]</sup> ..4. *DNA-experiments*

## Another example: general relativity

Einstein's theory of General Relativity makes several specific predictions about the observable structure of space-time, such as that light bends in a gravitational field, and that the amount of bending depends in a precise way on the strength of that gravitational field. Arthur Eddington's observations made during a 1919 solar eclipse supported General Relativity rather than Newtonian gravitation.<sup>[60]</sup>



Einstein's prediction (1907): Light bends in a gravitational field

## Experiments

Once predictions are made, they can be sought by experiments. If the test results contradict the predictions, the hypotheses which entailed them are called into question and become less tenable.

Sometimes the experiments are conducted incorrectly or are not very well designed, when compared to a crucial experiment. If the experimental results confirm the predictions, then the hypotheses are considered more likely to be correct, but might still be wrong and continue to be subject to further testing. The experimental control is a technique for dealing with observational error. This technique uses the contrast between multiple samples (or observations) under differing conditions to see what varies or what remains the same. We vary the conditions for each measurement, to help isolate what has changed. Mill's canons can then help us figure out what the important factor is.<sup>[61]</sup> Factor analysis is one technique for discovering the important factor in an effect.

Depending on the predictions, the experiments can have different shapes. It could be a classical experiment in a laboratory setting, a double-blind study or an archaeological excavation. Even taking a plane from New York to Paris is an experiment which tests the aerodynamical hypotheses used for constructing the plane.

Scientists assume an attitude of openness and accountability on the part of those conducting an experiment. Detailed record keeping is essential, to aid in recording and reporting on the experimental results, and supports the effectiveness and integrity of the procedure. They will also assist in reproducing the experimental results, likely by others. Traces of this approach can be seen in the work of Hipparchus (190-120 BCE), when determining a value for the precession of the Earth, while controlled experiments can be seen in the works of Jābir ibn Hayyān (721-815 CE), al-Battani (853–929) and Alhazen (965-1039).<sup>[62]</sup>

### DNA-experiments



Watson and Crick showed an initial (and incorrect) proposal for the structure of DNA to a team from Kings College - Rosalind Franklin, Maurice Wilkins, and Raymond Gosling. Franklin immediately spotted the flaws which concerned the water content. Later Watson saw Franklin's detailed X-ray diffraction images which showed an X-shape<sup>[63]</sup> and was able to confirm the structure was helical.<sup>[21][22]</sup>

This rekindled Watson and Crick's model building and led to the correct structure. ...1. *DNA-characterizations*

### Evaluation and improvement

The scientific method is iterative. At any stage it is possible to refine its accuracy and precision, so that some consideration will lead the scientist to repeat an earlier part of the process. Failure to develop an interesting hypothesis may lead a scientist to re-define the subject under consideration. Failure of a hypothesis to produce interesting and testable predictions may lead to reconsideration of the hypothesis or of the definition of the subject. Failure of an experiment to produce interesting results may lead a scientist to reconsider the experimental method, the hypothesis, or the definition of the subject.

Other scientists may start their own research and enter the process at any stage. They might adopt the characterization and formulate their own hypothesis, or they might adopt the hypothesis and deduce their own predictions. Often the experiment is not done by the person who made the prediction, and the characterization is based on experiments done by someone else. Published results of experiments can also serve as a hypothesis predicting their own reproducibility.

### DNA-iterations



After considerable fruitless experimentation, being discouraged by their superior from continuing, and numerous false starts,<sup>[64][65][66]</sup> Watson and Crick were able to infer the essential structure of DNA by concrete modeling of the physical shapes of the nucleotides which comprise it.<sup>[23][67]</sup> They were guided by the bond lengths which had been deduced by Linus Pauling and by Rosalind Franklin's X-ray

diffraction images. ...*DNA Example*

### Confirmation

Science is a social enterprise, and scientific work tends to be accepted by the scientific community when it has been confirmed. Crucially, experimental and theoretical results must be reproduced by others within the scientific community. Researchers have given their lives for this vision; Georg Wilhelm Richmann was killed by ball lightning (1753) when attempting to replicate the 1752 kite-flying experiment of Benjamin Franklin.<sup>[68]</sup>

To protect against bad science and fraudulent data, government research-granting agencies such as the National Science Foundation, and science journals including *Nature* and *Science*, have a policy that researchers must archive their data and methods so other researchers can test the data and methods and build on the research that has gone before. Scientific data archiving can be done at a number of national archives in the U.S. or in the World Data Center.

## Models of scientific inquiry

### Classical model

The classical model of scientific inquiry derives from Aristotle,<sup>[69]</sup> who distinguished the forms of approximate and exact reasoning, set out the threefold scheme of abductive, deductive, and inductive inference, and also treated the compound forms such as reasoning by analogy.

### Pragmatic model

In 1877,<sup>[70]</sup> Charles Sanders Peirce (pron.: /'pɜrs/ like "purse"; 1839–1914) characterized inquiry in general not as the pursuit of truth *per se* but as the struggle to move from irritating, inhibitory doubts born of surprises, disagreements, and the like, and to reach a secure belief, belief being that on which one is prepared to act. He framed scientific inquiry as part of a broader spectrum and as spurred, like inquiry generally, by actual doubt, not mere verbal or hyperbolic doubt, which he held to be fruitless.<sup>[71]</sup> He outlined four methods of settling opinion, ordered from least to most successful:

1. The method of tenacity (policy of sticking to initial belief) — which brings comforts and decisiveness but leads to trying to ignore contrary information and others' views as if truth were intrinsically private, not public. It goes against the social impulse and easily falters since one may well notice when another's opinion is as good as one's own initial opinion. Its successes can shine but tend to be transitory.
2. The method of authority — which overcomes disagreements but sometimes brutally. Its successes can be majestic and long-lived, but it cannot operate thoroughly enough to suppress doubts indefinitely, especially when people learn of other societies present and past.
3. The method of congruity or the *a priori* or the dilettante or "what is agreeable to reason" — which promotes conformity less brutally but depends on taste and fashion in paradigms and can go in circles over time, along with barren disputation. It is more intellectual and respectable but, like the first two methods, sustains accidental and capricious beliefs, destining some minds to doubts.
4. The scientific method — the method wherein inquiry regards itself as fallible and purposely tests itself and criticizes, corrects, and improves itself.

Peirce held that slow, stumbling ratiocination can be dangerously inferior to instinct and traditional sentiment in practical matters, and that the scientific method is best suited to theoretical research,<sup>[72]</sup> which in turn should not be trammelled by the other methods and practical ends; reason's "first rule" is that, in order to learn, one must desire to learn and, as a corollary, must not block the way of inquiry.<sup>[73]</sup> The scientific method excels the others by being deliberately designed to arrive — eventually — at the most secure beliefs, upon which the most successful practices can be based. Starting from the idea that people seek not truth *per se* but instead to subdue irritating, inhibitory doubt, Peirce showed how, through the struggle, some can come to submit to truth for the sake of belief's integrity, seek as truth the guidance of potential practice correctly to its given goal, and wed themselves to the scientific method.<sup>[70][74]</sup>

For Peirce, rational inquiry implies presuppositions about truth and the real; to reason is to presuppose (and at least to hope), as a principle of the reasoner's self-regulation, that the real is discoverable and independent of our vagaries of opinion. In that vein he defined truth as the correspondence of a sign (in particular, a proposition) to its object and, pragmatically, not as actual consensus of some definite, finite community (such that to inquire would be to poll the experts), but instead as that final opinion which all investigators *would* reach sooner or later but still inevitably, if they were to push investigation far enough, even when they start from different points.<sup>[75]</sup> In tandem he defined the real as a true sign's object (be that object a possibility or quality, or an actuality or brute fact, or a necessity or norm or law), which is what it is independently of any finite community's opinion and, pragmatically, depends only on the final opinion destined in a sufficient investigation. That is a destination as far, or near, as the truth itself to you or me or the given finite community. Thus his theory of inquiry boils down to "Do the science." Those conceptions of truth

and the real involve the idea of a community both without definite limits (and thus potentially self-correcting as far as needed) and capable of definite increase of knowledge.<sup>[76]</sup> As inference, "logic is rooted in the social principle" since it depends on a standpoint that is, in a sense, unlimited.<sup>[77]</sup>

Paying special attention to the generation of explanations, Peirce outlined the scientific method as a coordination of three kinds of inference in a purposeful cycle aimed at settling doubts, as follows (in §III–IV in "A Neglected Argument"<sup>[78]</sup> except as otherwise noted):

1. **Abduction** (or retrodution). Guessing, inference to explanatory hypotheses for selection of those best worth trying. From abduction, Peirce distinguishes induction as inferring, on the basis of tests, the proportion of truth in the hypothesis. Every inquiry, whether into ideas, brute facts, or norms and laws, arises from surprising observations in one or more of those realms (and for example at any stage of an inquiry already underway). All explanatory content of theories comes from abduction, which guesses a new or outside idea so as to account in a simple, economical way for a surprising or complicative phenomenon. Oftenest, even a well-prepared mind guesses wrong. But the modicum of success of our guesses far exceeds that of sheer luck and seems born of attunement to nature by instincts developed or inherent, especially insofar as best guesses are optimally plausible and simple in the sense, said Peirce, of the "facile and natural", as by Galileo's natural light of reason and as distinct from "logical simplicity". Abduction is the most fertile but least secure mode of inference. Its general rationale is inductive: it succeeds often enough and, without it, there is no hope of sufficiently expediting inquiry (often multi-generational) toward new truths.<sup>[79]</sup> Coordinative method leads from abducting a plausible hypothesis to judging it for its testability<sup>[80]</sup> and for how its trial would economize inquiry itself.<sup>[81]</sup> Peirce calls his pragmatism "the logic of abduction".<sup>[82]</sup> His pragmatic maxim is: "Consider what effects that might conceivably have practical bearings you conceive the objects of your conception to have. Then, your conception of those effects is the whole of your conception of the object".<sup>[75]</sup> His pragmatism is a method of reducing conceptual confusions fruitfully by equating the meaning of any conception with the conceivable practical implications of its object's conceived effects — a method of experimental mental reflection hospitable to forming hypotheses and conducive to testing them. It favors efficiency. The hypothesis, being insecure, needs to have practical implications leading at least to mental tests and, in science, lending themselves to scientific tests. A simple but unlikely guess, if uncostly to test for falsity, may belong first in line for testing. A guess is intrinsically worth testing if it has instinctive plausibility or reasoned objective probability, while subjective likelihood, though reasoned, can be misleadingly seductive. Guesses can be chosen for trial strategically, for their caution (for which Peirce gave as example the game of Twenty Questions), breadth, and incompleteness.<sup>[83]</sup> One can hope to discover only that which time would reveal through a learner's sufficient experience anyway, so the point is to expedite it; the economy of research is what demands the leap, so to speak, of abduction and governs its art.<sup>[81]</sup>

2. **Deduction**. Two stages:

- i. Explication. Unclearly premissed, but deductive, analysis of the hypothesis in order to render its parts as clear as possible.
- ii. Demonstration: Deductive Argumentation, Euclidean in procedure. Explicit deduction of hypothesis's consequences as predictions, for induction to test, about evidence to be found. Corollarial or, if needed, Theorematic.

3. **Induction**. The long-run validity of the rule of induction is deducible from the principle (presuppositional to reasoning in general<sup>[75]</sup>) that the real is only the object of the final opinion to which adequate investigation would lead;<sup>[84]</sup> anything to which no such process would ever lead would not be real. Induction involving ongoing tests or observations follows a method which, sufficiently persisted in, will diminish its error below any predesignate degree. Three stages:

- i. Classification. Unclearly premissed, but inductive, classing of objects of experience under general ideas.
- ii. Probation: direct Inductive Argumentation. Crude (the enumeration of instances) or Gradual (new estimate of proportion of truth in the hypothesis after each test). Gradual Induction is Qualitative or Quantitative; if

Qualitative, then dependent on weightings of qualities or characters;<sup>[85]</sup> if Quantitative, then dependent on measurements, or on statistics, or on countings.

iii. Sentential Induction. "...which, by Inductive reasonings, appraises the different Probations singly, then their combinations, then makes self-appraisal of these very appraisals themselves, and passes final judgment on the whole result".

### **Computational approaches**

Many subspecialties of applied logic and computer science, such as artificial intelligence, machine learning, computational learning theory, inferential statistics, and knowledge representation, are concerned with setting out computational, logical, and statistical frameworks for the various types of inference involved in scientific inquiry. In particular, they contribute hypothesis formation, logical deduction, and empirical testing. Some of these applications draw on measures of complexity from algorithmic information theory to guide the making of predictions from prior distributions of experience, for example, see the complexity measure called the *speed prior* from which a computable strategy for optimal inductive reasoning can be derived.

### **Communication and community**

Frequently the scientific method is employed not only by a single person, but also by several people cooperating directly or indirectly. Such cooperation can be regarded as one of the defining elements of a scientific community. Various techniques have been developed to ensure the integrity of scientific methodology within such an environment.

### **Peer review evaluation**

Scientific journals use a process of *peer review*, in which scientists' manuscripts are submitted by editors of scientific journals to (usually one to three) fellow (usually anonymous) scientists familiar with the field for evaluation. The referees may or may not recommend publication, publication with suggested modifications, or, sometimes, publication in another journal. This serves to keep the scientific literature free of unscientific or pseudoscientific work, to help cut down on obvious errors, and generally otherwise to improve the quality of the material. The peer review process can have limitations when considering research outside the conventional scientific paradigm: problems of "groupthink" can interfere with open and fair deliberation of some new research.<sup>[86]</sup>

### **Documentation and replication**

Sometimes experimenters may make systematic errors during their experiments, unconsciously veer from scientific method (Pathological science) for various reasons, or, in rare cases, deliberately report false results. Consequently, it is a common practice for other scientists to attempt to repeat the experiments in order to duplicate the results, thus further validating the hypothesis.

### **Archiving**

As a result, researchers are expected to practice scientific data archiving in compliance with the policies of government funding agencies and scientific journals. Detailed records of their experimental procedures, raw data, statistical analyses and source code are preserved in order to provide evidence of the effectiveness and integrity of the procedure and assist in reproduction. These procedural records may also assist in the conception of new experiments to test the hypothesis, and may prove useful to engineers who might examine the potential practical applications of a discovery.



### Data sharing

When additional information is needed before a study can be reproduced, the author of the study is expected to provide it promptly. If the author refuses to share data, appeals can be made to the journal editors who published the study or to the institution which funded the research.

### Limitations

Since it is impossible for a scientist to record *everything* that took place in an experiment, facts selected for their apparent relevance are reported. This may lead, unavoidably, to problems later if some supposedly irrelevant feature is questioned. For example, Heinrich Hertz did not report the size of the room used to test Maxwell's equations, which later turned out to account for a small deviation in the results. The problem is that parts of the theory itself need to be assumed in order to select and report the experimental conditions. The observations are hence sometimes described as being 'theory-laden'.

### Dimensions of practice

The primary constraints on contemporary science are:

- Publication, i.e. Peer review
- Resources (mostly funding)

It has not always been like this: in the old days of the "gentleman scientist" funding (and to a lesser extent publication) were far weaker constraints.

Both of these constraints indirectly require scientific method — work that violates the constraints will be difficult to publish and difficult to get funded. Journals require submitted papers to conform to "good scientific practice" and this is mostly enforced by peer review. Originality, importance and interest are more important - see for example the author guidelines<sup>[87]</sup> for *Nature*.

### Philosophy and sociology of science

Philosophy of science looks at the underpinning logic of the scientific method, at what separates science from non-science, and the ethic that is implicit in science. There are basic assumptions derived from philosophy that form the base of the scientific method - namely, that reality is objective and consistent, that humans have the capacity to perceive reality accurately, and that rational explanations exist for elements of the real world. These assumptions from methodological naturalism form the basis on which science is grounded. Logical Positivist, empiricist, falsificationist, and other theories have claimed to give a definitive account of the logic of science, but each has in turn been criticized.

Thomas Kuhn examined the history of science in his *The Structure of Scientific Revolutions*, and found that the actual method used by scientists differed dramatically from the then-espoused method. His observations of science practice are essentially sociological and do not speak to how science is or can be practiced in other times and other cultures.

Norwood Russell Hanson, Imre Lakatos and Thomas Kuhn have done extensive work on the "theory laden" character of observation. Hanson (1958) first coined the term for the idea that all observation is dependent on the conceptual framework of the observer, using the concept of gestalt to show how preconceptions can affect both observation and description.<sup>[88]</sup> He opens Chapter 1 with a discussion of the Golgi bodies and their initial rejection as an artefact of staining technique, and a discussion of Brahe and Kepler observing the dawn and seeing a "different" sun rise despite the same physiological phenomenon. Kuhn<sup>[89]</sup> and Feyerabend<sup>[90]</sup> acknowledge the pioneering significance of his work.

Kuhn (1961) said the scientist generally has a theory in mind before designing and undertaking experiments so as to make empirical observations, and that the "route from theory to measurement can almost never be traveled

backward". This implies that the way in which theory is tested is dictated by the nature of the theory itself, which led Kuhn (1961, p. 166) to argue that "once it has been adopted by a profession ... no theory is recognized to be testable by any quantitative tests that it has not already passed".<sup>[91]</sup>

Paul Feyerabend similarly examined the history of science, and was led to deny that science is genuinely a methodological process. In his book *Against Method* he argues that scientific progress is *not* the result of applying any particular method. In essence, he says that for any specific method or norm of science, one can find a historic episode where violating it has contributed to the progress of science. Thus, if believers in scientific method wish to express a single universally valid rule, Feyerabend jokingly suggests, it should be 'anything goes'.<sup>[92]</sup> Criticisms such as his led to the strong programme, a radical approach to the sociology of science.

The postmodernist critiques of science have themselves been the subject of intense controversy. This ongoing debate, known as the science wars, is the result of conflicting values and assumptions between the postmodernist and realist camps. Whereas postmodernists assert that scientific knowledge is simply another discourse (note that this term has special meaning in this context) and not representative of any form of fundamental truth, realists in the scientific community maintain that scientific knowledge does reveal real and fundamental truths about reality. Many books have been written by scientists which take on this problem and challenge the assertions of the postmodernists while defending science as a legitimate method of deriving truth.<sup>[93]</sup>

### **Role of chance in discovery**

Somewhere between 33% and 50% of all scientific discoveries are estimated to have been *stumbled upon*, rather than sought out. This may explain why scientists so often express that they were lucky.<sup>[94]</sup> Louis Pasteur is credited with the famous saying that "Luck favours the prepared mind", but some psychologists have begun to study what it means to be 'prepared for luck' in the scientific context. Research is showing that scientists are taught various heuristics that tend to harness chance and the unexpected.<sup>[94][95]</sup> This is what professor of economics Nassim Nicholas Taleb calls "Anti-fragility"; while some systems of investigation are fragile in the face of human error, human bias, and randomness, the scientific method is more than resistant or tough - it actually benefits from such randomness in many ways (it is anti-fragile). Taleb believes that the more anti-fragile the system, the more it will flourish in the real world.<sup>[96]</sup>

Psychologist Kevin Dunbar says the process of discovery often starts with researchers finding bugs in their experiments. These unexpected results lead researchers to try and fix what they *think* is an error in their method. Eventually, the researcher decides the error is too persistent and systematic to be a coincidence. The highly controlled, cautious and curious aspects of the scientific method are thus what make it well suited for identifying such persistent systematic errors. At this point, the researcher will begin to think of theoretical explanations for the error, often seeking the help of colleagues across different domains of expertise.<sup>[94][95]</sup>

## History

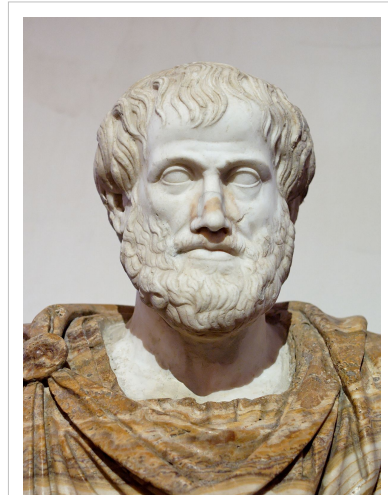
The development of the scientific method is inseparable from the history of science itself. Ancient Egyptian documents describe empirical methods in astronomy,<sup>[98]</sup> mathematics,<sup>[99]</sup> and medicine.<sup>[100]</sup> The ancient Greek philosopher Thales in the 6th century BC refused to accept supernatural, religious or mythological explanations for natural phenomena, proclaiming that every event had a natural cause. The development of deductive reasoning by Plato was an important step towards the scientific method. Empiricism seems to have been formalized by Aristotle, who believed that universal truths could be reached via induction.

There are hints of experimental methods from the Classical world (e.g., those reported by Archimedes in a report recovered early in the 20th century from an overwritten manuscript), but the first clear instances of an experimental scientific method seem to have been developed by Islamic scientists who introduced the use of experimentation and quantification within a generally empirical orientation. For example, Alhazen performed optical and physiological experiments, reported in his manifold works, the most famous being *Book of Optics* (1021).<sup>[101]</sup>

By the late 15th century, the physician-scholar Niccolò Leonicensi was finding errors in Pliny's *Natural History*. As a physician, Leonicensi was concerned about these botanical errors propagating to the materia medica on which medicines were based.<sup>[102]</sup> To counter this, a botanical garden was established at Orto botanico di Padova, University of Padua (in use for teaching by 1546), in order that medical students might have empirical access to the plants of a pharmacopia. The philosopher and physician Francisco Sanchez was led by his medical training at Rome, 1571–73, and by the philosophical skepticism recently placed in the European mainstream by the publication of Sextus Empiricus' "Outlines of Pyrrhonism", to search for a true method of knowing (*modus sciendi*), as nothing clear can be known by the methods of Aristotle and his followers<sup>[103]</sup> — for example, syllogism fails upon circular reasoning. Following the physician Galen's *method of medicine*, Sanchez lists the methods of judgement and experience, which are faulty in the wrong hands,<sup>[104]</sup> and we are left with the bleak statement *That Nothing is Known* (1581). This challenge was taken up by René Descartes in the next generation (1637), but at the least, Sanchez warns us that we ought to refrain from the methods, summaries, and commentaries on Aristotle, if we seek scientific knowledge. In this, he is echoed by Francis Bacon, also influenced by skepticism; Sanchez cites the humanist Juan Luis Vives who sought a better educational system, as well as a statement of human rights as a pathway for improvement of the lot of the poor.

The modern scientific method crystallized no later than in the 17th and 18th centuries. In his work *Novum Organum* (1620) — a reference to Aristotle's *Organon* — Francis Bacon outlined a new system of logic to improve upon the old philosophical process of syllogism.<sup>[105]</sup> Then, in 1637, René Descartes established the framework for scientific method's guiding principles in his treatise, *Discourse on Method*. The writings of Alhazen, Bacon and Descartes are considered critical in the historical development of the modern scientific method, as are those of John Stuart Mill.<sup>[106]</sup>

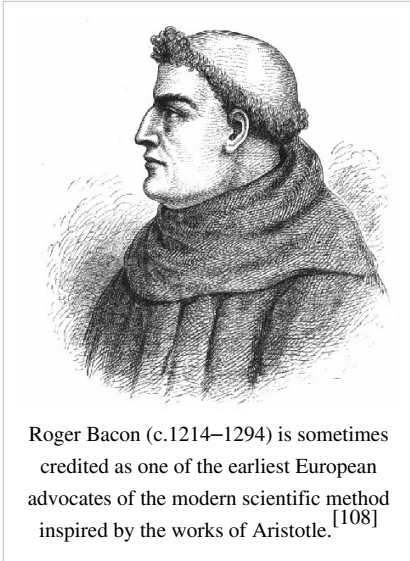
Grosseteste was "the principal figure" in bringing about "a more adequate method of scientific inquiry" by which "medieval scientists were able eventually to outstrip their ancient European and Muslim teachers" (Dales 1973:62). ... His thinking influenced Roger Bacon, who spread Grosseteste's ideas from Oxford to the University of Paris during a visit there in the 1240s. From the prestigious universities in Oxford and Paris, the new experimental science spread rapidly throughout the medieval universities: "And so it went to Galileo,



Aristotle, 384 BC–322 BC. "As regards his method, Aristotle is recognized as the inventor of scientific method because of his refined analysis of logical implications contained in demonstrative discourse, which goes well beyond natural logic and does not owe anything to the ones who philosophized before him."—Riccardo Pozzo<sup>[97]</sup>

William Gilbert, Francis Bacon, William Harvey, Descartes, Robert Hooke, Newton, Leibniz, and the world of the seventeenth century" (Crombie 1962:15). So it went to us also.

— Hugh G. Gauch, 2003.<sup>[107]</sup>



Roger Bacon (c.1214–1294) is sometimes credited as one of the earliest European advocates of the modern scientific method inspired by the works of Aristotle.<sup>[108]</sup>

In the late 19th century, Charles Sanders Peirce proposed a schema that would turn out to have considerable influence in the development of current scientific methodology generally. Peirce accelerated the progress on several fronts. Firstly, speaking in broader context in "How to Make Our Ideas Clear" (1878)<sup>[109]</sup>, Peirce outlined an objectively verifiable method to test the truth of putative knowledge on a way that goes beyond mere foundational alternatives, focusing upon both *deduction* and *induction*. He thus placed induction and deduction in a complementary rather than competitive context (the latter of which had been the primary trend at least since David Hume, who wrote in the mid-to-late 18th century). Secondly, and of more direct importance to modern method, Peirce put forth the basic schema for hypothesis/testing that continues to prevail today. Extracting the theory of inquiry from its raw materials in classical logic, he refined it in parallel with the early development of symbolic logic to address the then-current problems in scientific reasoning. Peirce examined and

articulated the three fundamental modes of reasoning that, as discussed above in this article, play a role in inquiry today, the processes that are currently known as abductive, deductive, and inductive inference. Thirdly, he played a major role in the progress of symbolic logic itself — indeed this was his primary specialty.

Beginning in the 1930s, Karl Popper argued that there is no such thing as inductive reasoning.<sup>[110]</sup> All inferences ever made, including in science, are purely<sup>[111]</sup> deductive according to this view. Accordingly, he claimed that the empirical character of science has nothing to do with induction—but with the deductive property of falsifiability that scientific hypotheses have. Contrasting his views with inductivism and positivism, he even denied the existence of the scientific method: "(1) There is no method of discovering a scientific theory (2) There is no method for ascertaining the truth of a scientific hypothesis, i.e., no method of verification; (3) There is no method for ascertaining whether a hypothesis is 'probable', or probably true".<sup>[112]</sup> Instead, he held that there is only one universal method, a method not particular to science: The negative method of criticism, or colloquially termed trial and error. It covers not only all products of the human mind, including science, mathematics, philosophy, art and so on, but also the evolution of life. Following Peirce and others, Popper argued that science is fallible and has no authority.<sup>[112]</sup> In contrast to empiricist-inductivist views, he welcomed metaphysics and philosophical discussion and even gave qualified support to myths<sup>[113]</sup> and pseudosciences.<sup>[114]</sup> Popper's view has become known as critical rationalism.

## Relationship with mathematics

Science is the process of gathering, comparing, and evaluating proposed models against observables. A model can be a simulation, mathematical or chemical formula, or set of proposed steps. Science is like mathematics in that researchers in both disciplines can clearly distinguish what is *known* from what is *unknown* at each stage of discovery. Models, in both science and mathematics, need to be internally consistent and also ought to be *falsifiable* (capable of disproof). In mathematics, a statement need not yet be proven; at such a stage, that statement would be called a conjecture. But when a statement has attained mathematical proof, that statement gains a kind of immortality which is highly prized by mathematicians, and for which some mathematicians devote their lives.<sup>[115]</sup>

Mathematical work and scientific work can inspire each other.<sup>[116]</sup> For example, the technical concept of time arose in science, and timelessness was a hallmark of a mathematical topic. But today, the Poincaré conjecture has been

proven using time as a mathematical concept in which objects can flow (see Ricci flow).

Nevertheless, the connection between mathematics and reality (and so science to the extent it describes reality) remains obscure. Eugene Wigner's paper, *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*, is a very well known account of the issue from a Nobel Prize physicist. In fact, some observers (including some well known mathematicians such as Gregory Chaitin, and others such as Lakoff and Núñez) have suggested that mathematics is the result of practitioner bias and human limitation (including cultural ones), somewhat like the post-modernist view of science.

George Pólya's work on problem solving,<sup>[117]</sup> the construction of mathematical proofs, and heuristic<sup>[118][119]</sup> show that the mathematical method and the scientific method differ in detail, while nevertheless resembling each other in using iterative or recursive steps.

	Mathematical method	Scientific method
1	Understanding	Characterization from experience and observation
2	Analysis	Hypothesis: a proposed explanation
3	Synthesis	Deduction: prediction from the hypothesis
4	Review/Extend	Test and experiment

In Pólya's view, *understanding* involves restating unfamiliar definitions in your own words, resorting to geometrical figures, and questioning what we know and do not know already; *analysis*, which Pólya takes from Pappus,<sup>[120]</sup> involves free and heuristic construction of plausible arguments, working backward from the goal, and devising a plan for constructing the proof; *synthesis* is the strict Euclidean exposition of step-by-step details<sup>[121]</sup> of the proof; *review* involves reconsidering and re-examining the result and the path taken to it.

Gauss, when asked how he came about his theorems, once replied "durch planmässiges Tattonieren" (through systematic palpable experimentation).<sup>[122]</sup>

Imre Lakatos argued that mathematicians actually use contradiction, criticism and revision as principles for improving their work.<sup>[123]</sup>

## Notes

[1] Goldhaber & Nieto 2010, p. 940

[2] "[4] Rules for the study of natural philosophy", Newton 1999, pp. 794–6, from Book 3, *The System of the World*.

[3] Oxford English Dictionary - entry for *scientific*.

[4] "How does light travel through transparent bodies? Light travels through transparent bodies in straight lines only.... We have explained this exhaustively in our *Book of Optics*. But let us now mention something to prove this convincingly: the fact that light travels in straight lines is clearly observed in the lights which enter into dark rooms through holes.... [T]he entering light will be clearly observable in the dust which fills the air. —Alhazen, translated into English from German by M. Schwarz, from "Abhandlung über das Licht", J. Baermann (ed. 1882) *Zeitschrift der Deutschen Morgenländischen Gesellschaft* Vol 36 as quoted in Sambursky 1974, p. 136.

- He demonstrated his conjecture that "light travels through transparent bodies in straight lines only" by placing a straight stick or a taut thread next to the light beam, as quoted in Sambursky 1974, p. 136 to prove that light travels in a straight line.
- David Hockney, (2001, 2006) in *Secret Knowledge: rediscovering the lost techniques of the old masters* ISBN 0-14-200512-6 (expanded edition) cites Alhazen several times as the likely source for the portraiture technique using the camera obscura, which Hockney rediscovered with the aid of an optical suggestion from Charles M. Falco. *Kitab al-Manazir*, which is Alhazen's *Book of Optics*, at that time denoted *Opticae Thesaurus, Alhazen Arabis*, was translated from Arabic into Latin for European use as early as 1270. Hockney cites Friedrich Risner's 1572 Basle edition of *Opticae Thesaurus*. Hockney quotes Alhazen as the first clear description of the camera obscura in Hockney, p. 240.

"Truth is sought for its own sake. And those who are engaged upon the quest for anything for its own sake are not interested in other things. Finding the truth is difficult, and the road to it is rough."—Alhazen (Ibn Al-Haytham 965-c.1040) *Critique of Ptolemy*, translated by S. Pines, *Actes X Congrès internationale d'histoire des sciences*, Vol I Ithaca 1962, as quoted in Sambursky 1974, p. 139. (This quotation is from Alhazen's critique of Ptolemy's books *Almagest*, *Planetary Hypotheses*, and *Optics* as translated into English by A. Mark Smith ([http://books.google.com/books?id=mhLVHR5QAQkC&pg=PA59&lpg=PA59&dq=Opticae+thesaurus+alhazen&source=bl&ots=noo2fzmnU-&sig=fHI2OZUVkiKuxyOGw-nt08p9ISM&hl=en&ei=QHU1TZnCsT68AbywuTyCA&sa=X&oi=book\\_result&ct=result&resnum=1&](http://books.google.com/books?id=mhLVHR5QAQkC&pg=PA59&lpg=PA59&dq=Opticae+thesaurus+alhazen&source=bl&ots=noo2fzmnU-&sig=fHI2OZUVkiKuxyOGw-nt08p9ISM&hl=en&ei=QHU1TZnCsT68AbywuTyCA&sa=X&oi=book_result&ct=result&resnum=1&)).

- ved=0CBIQ6AEwADgK#v=onepage&q=Opticae thesaurus alhazen&f=false.)
- [5] Morris Kline (1985) *Mathematics for the nonmathematician* (<http://books.google.com/books?id=f-e0bro-0FUC&pg=PA284&dq&hl=en#v=onepage&q=&f=false>). Courier Dover Publications. p. 284. ISBN 0-486-24823-2
- [6] Shapere, Dudley (1974). *Galileo: A Philosophical Study*. University of Chicago Press. ISBN 0-226-75007-8.
- [7] Peirce, C. S., *Collected Papers* v. 1, paragraph 74.
- [8] Karl R. Popper (1963), 'The Logic of Scientific Discovery'. The Logic of Scientific Discovery (<http://www.cosmopolitanuniversity.ac/library/LogicofScientificDiscoveryPopper1959.pdf>) pp. 17-20, 249-252, 437-438, and elsewhere.
- Leon Lederman, for teaching physics first, illustrates how to avoid confirmation bias: Ian Shelton, in Chile, was initially skeptical that supernova 1987a was real, but possibly an artifact of instrumentation (null hypothesis), so he went outside and disproved his null hypothesis by observing SN 1987a with the naked eye. The Kamiokande experiment, in Japan, independently observed neutrinos from SN 1987a at the same time.
- [9] Gauch 2003, pp. 3
- [10] *History of Inductive Science* (1837), and in *Philosophy of Inductive Science* (1840)
- [11] Schuster and Powers (2005), Translational and Experimental Clinical Research, Ch. 1. Link. ([http://books.google.com/books?id=C7pZftbI0ZMC&printsec=frontcover&dq=Translational+and+Experimental+Clinical+Research&hl=en&sa=X&ei=ZciuT\\_SzCOnOiAKNzIzfAw&ved=0CE4Q6AEwAA#v=onepage&q=Translational and Experimental Clinical Research&f=false](http://books.google.com/books?id=C7pZftbI0ZMC&printsec=frontcover&dq=Translational+and+Experimental+Clinical+Research&hl=en&sa=X&ei=ZciuT_SzCOnOiAKNzIzfAw&ved=0CE4Q6AEwAA#v=onepage&q=Translational and Experimental Clinical Research&f=false)) This chapter also discusses the different types of research questions and how they are produced.
- [12] This phrasing is attributed to Marshall Nirenberg.
- [13] Karl R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge, 2003 ISBN 0-415-28594-1
- [14] Galilei, Galileo (M.D.C.XXXVIII), *Discorsi e Dimostrazioni Matematiche, intorno a due nuoue scienze*, Leida: Apresso gli Elsevirri, ISBN 0-486-60099-8, Dover reprint of the 1914 Macmillan translation by Henry Crew and Alfonso de Salvio of *Two New Sciences*, Galileo Galilei Linceo (1638). Additional publication information is from the collection of first editions of the Library of Congress surveyed by Bruno 1989, pp. 261–264.
- [15] Godfrey-Smith 2003 p. 236.
- [16] McCarty 1985
- [17] October 1951, as noted in McElheny 2004, p. 40: "That's what a helix should look like!" Crick exclaimed in delight (This is the Cochran-Crick-Vand-Stokes theory of the transform of a helix).
- [18] June 1952, as noted in McElheny 2004, p. 43: Watson had succeeded in getting X-ray pictures of TMV showing a diffraction pattern consistent with the transform of a helix.
- [19] Watson did enough work on Tobacco mosaic virus to produce the diffraction pattern for a helix, per Crick's work on the transform of a helix. pp. 137-138, Horace Freeland Judson (1979) *The Eighth Day of Creation* ISBN 0-671-22540-5
- [20] — Cochran W, Crick FHC and Vand V. (1952) "The Structure of Synthetic Polypeptides. I. The Transform of Atoms on a Helix", *Acta Cryst.*, **5**, 581-586.
- [21] Friday, January 30, 1953. Tea time, as noted in McElheny 2004, p. 52: Franklin confronts Watson and his paper - "Of course it [Pauling's pre-print] is wrong. DNA is not a helix." However, Watson then visits Wilkins' office, sees photo 51, and immediately recognizes the diffraction pattern of a helical structure. But additional questions remained, requiring additional iterations of their research. For example, the number of strands in the backbone of the helix (Crick suspected 2 strands, but cautioned Watson to examine that more critically), the location of the base pairs (inside the backbone or outside the backbone), etc. One key point was that they realized that the quickest way to reach a result was not to continue a mathematical analysis, but to build a physical model.
- [22] "The instant I saw the picture my mouth fell open and my pulse began to race." —Watson 1968, p. 167 Page 168 shows the X-shaped pattern of the B-form of DNA, clearly indicating crucial details of its helical structure to Watson and Crick.
- McElheny 2004 p.52 dates the Franklin-Watson confrontation as Friday, January 30, 1953. Later that evening, Watson urges Wilkins to begin model-building immediately. But Wilkins agrees to do so only after Franklin's departure.
- [23] Saturday, February 28, 1953, as noted in McElheny 2004, pp. 57–59: Watson found the base pairing mechanism which explained Chargaff's rules using his cardboard models.
- [24] Fleck 1979, pp. xxvii-xxviii
- [25] " NIH Data Sharing Policy ([http://grants.nih.gov/grants/policy/data\\_sharing/index.htm](http://grants.nih.gov/grants/policy/data_sharing/index.htm))."
- [26] Stanovich, Keith E. (2007). *How to Think Straight About Psychology*. Boston: Pearson Education. pg 123
- [27] Brody 1993, pp. 44–45
- [28] Hall, B. K.; Hallgrímsson, B., eds. (2008). *Strickberger's Evolution* (<http://www.jblearning.com/catalog/9780763700669/>) (4th ed.). Jones & Bartlett. p. 762. ISBN 0-7637-0066-5. .
- [29] Cracraft, J.; Donoghue, M. J., eds. (2005). *Assembling the tree of life* ([http://books.google.ca/books?id=6IXTP0YU6\\_kC&printsec=frontcover&dq=Assembling+the+tree+of+life#v=onepage&q&f=false](http://books.google.ca/books?id=6IXTP0YU6_kC&printsec=frontcover&dq=Assembling+the+tree+of+life#v=onepage&q&f=false)). Oxford University Press. p. 592. ISBN 0-19-517234-5. .
- [30] Needham & Wang 1954 p.166 shows how the 'flying gallop' image propagated from China to the West.
- [31] "A myth is a belief given uncritical acceptance by members of a group ..." —Weiss, *Business Ethics* p. 15, as cited by Ronald R. Sims (2003) *Ethics and corporate social responsibility: why giants fall* p.21
- [32] Imre Lakatos (1976), *Proofs and Refutations*. Taleb 2007, p. 72 lists ways to avoid narrative fallacy and confirmation bias.
- [33] For more on the narrative fallacy, see also Fleck 1979, p. 27: "Words and ideas are originally phonetic and mental equivalences of the experiences coinciding with them. ... Such proto-ideas are at first always too broad and insufficiently specialized. ... Once a structurally

- complete and closed system of opinions consisting of many details and relations has been formed, it offers enduring resistance to anything that contradicts it."
- [34] See the hypothethico-deductive method, for example, Godfrey-Smith 2003, p. 236.
- [35] Jevons 1874, pp. 265–6.
- [36] pp.65,73,92,398 —Andrew J. Galambos, *Sic Itur ad Astra* ISBN 0-88078-004-5(AJG learned scientific method from Felix Ehrenhaft
- [37] Galileo 1638, pp. v-xii,1–300
- [38] Brody 1993, pp. 10–24 calls this the "epistemic cycle": "The epistemic cycle starts from an initial model; iterations of the cycle then improve the model until an adequate fit is achieved."
- [39] Iteration example: Chaldean astronomers such as Kidinnu compiled astronomical data. Hipparchus was to use this data to calculate the precession of the Earth's axis. Fifteen hundred years after Kidinnu, Al-Batani, born in what is now Turkey, would use the collected data and improve Hipparchus' value for the precession of the Earth's axis. Al-Batani's value, 54.5 arc-seconds per year, compares well to the current value of 49.8 arc-seconds per year (26,000 years for Earth's axis to round the circle of nutation).
- [40] Recursion example: the Earth is itself a magnet, with its own North and South Poles William Gilbert (in Latin 1600) *De Magnete, or On Magnetism and Magnetic Bodies*. Translated from Latin to English, selection by Moulton & Schifferes 1960, pp. 113–117. Gilbert created a *terrella*, a lodestone ground into a spherical shape, which served as Gilbert's model for the Earth itself, as noted in Bruno 1989, p. 277.
- [41] "The foundation of general physics ... is experience. These ... everyday experiences we do not discover without deliberately directing our attention to them. Collecting information about these is *observation*." —Hans Christian Ørsted("First Introduction to General Physics" ¶13, part of a series of public lectures at the University of Copenhagen. Copenhagen 1811, in Danish, printed by Johan Frederik Schulz. In Kirstine Meyer's 1920 edition of Ørsted's works, vol.III pp. 151-190. ) "First Introduction to Physics: the Spirit, Meaning, and Goal of Natural Science". Reprinted in German in 1822, Schweigger's *Journal für Chemie und Physik* 36, pp.458-488, as translated in Ørsted 1997, p. 292
- [42] "When it is not clear under which law of nature an effect or class of effect belongs, we try to fill this gap by means of a guess. Such guesses have been given the name *conjectures* or *hypotheses*." —Hans Christian Ørsted(1811) "First Introduction to General Physics" as translated in Ørsted 1997, p. 297.
- [43] "In general we look for a new law by the following process. First we guess it. ...", —Feynman 1965, p. 156
- [44] "... the statement of a law - A depends on B - always transcends experience."—Born 1949, p. 6
- [45] "The student of nature ... regards as his property the experiences which the mathematician can only borrow. This is why he deduces theorems directly from the nature of an effect while the mathematician only arrives at them circuitously." —Hans Christian Ørsted(1811) "First Introduction to General Physics" ¶17. as translated in Ørsted 1997, p. 297.
- [46] Salviati speaks: "I greatly doubt that Aristotle ever tested by experiment whether it be true that two stones, one weighing ten times as much as the other, if allowed to fall, at the same instant, from a height of, say, 100 cubits, would so differ in speed that when the heavier had reached the ground, the other would not have fallen more than 10 cubits." Two New Sciences (1638) (<http://galileo.phys.virginia.edu/classes/109N/tns61.htm>) —Galileo 1638, pp. 61–62. A more extended quotation is referenced by Moulton & Schifferes 1960, pp. 80–81.
- [47] In the inquiry-based education paradigm, the stage of "characterization, observation, definition, ..." is more briefly summed up under the rubric of a Question
- [48] "To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science." —Einstein & Infeld 1938, p. 92.
- [49] Crawford S, Stucki L (1990), "Peer review and the changing research record", "J Am Soc Info Science", vol. 41, pp 223-228
- [50] See, e.g., Gauch 2003, esp. chapters 5-8
- [51] Cartwright, Nancy (1983), *How the Laws of Physics Lie*. Oxford: Oxford University Press. ISBN 0-19-824704-4
- [52] Andreas Vesalius, *Epistola, Rationem, Modumque Propinandi Radicis Chynae Decocti* (1546), 141. Quoted and translated in C.D. O'Malley, *Andreas Vesalius of Brussels*, (1964), 116. As quoted by Bynum & Porter 2005, p. 597: Andreas Vesalius,597#1.
- [53] Crick, Francis (1994), *The Astonishing Hypothesis* ISBN 0-684-19431-7 p.20
- [54] McElheny 2004 p.34
- [55] Glen 1994, pp. 37–38.
- [56] "The structure that we propose is a three-chain structure, each chain being a helix" — Linus Pauling, as quoted on p. 157 by Horace Freeland Judson (1979), *The Eighth Day of Creation* ISBN 0-671-22540-5
- [57] McElheny 2004, pp. 49–50: January 28, 1953 - Watson read Pauling's pre-print, and realized that in Pauling's model, DNA's phosphate groups had to be un-ionized. But DNA is an acid, which contradicts Pauling's model.
- [58] June 1952. as noted in McElheny 2004, p. 43: Watson had succeeded in getting X-ray pictures of TMV showing a diffraction pattern consistent with the transform of a helix.
- [59] McElheny 2004 p.68: *Nature* April 25, 1953.
- [60] In March 1917, the Royal Astronomical Society announced that on May 29, 1919, the occasion of a total eclipse of the sun would afford favorable conditions for testing Einstein's General theory of relativity. One expedition, to Sobral, Ceará, Brazil, and Eddington's expedition to the island of Principe yielded a set of photographs, which, when compared to photographs taken at Sobral and at Greenwich Observatory showed that the deviation of light was measured to be 1.69 arc-seconds, as compared to Einstein's desk prediction of 1.75 arc-seconds. — Antonina Vallentin (1954), *Einstein*, as quoted by Samuel Rapport and Helen Wright (1965), *Physics*, New York: Washington Square Press, pp 294-295.
- [61] Mill, John Stuart, "A System of Logic", University Press of the Pacific, Honolulu, 2002, ISBN 1-4102-0252-6.



- [62] al-Battani, *De Motu Stellarum* translation from Arabic to Latin in 1116, as cited by "Battani, al-" (c.858-929) *Encyclopaedia Britannica*, 15th. ed. Al-Battani is known for his accurate observations at al-Raqqa in Syria, beginning in 877. His work includes measurement of the annual precession of the equinoxes.
- [63] <http://www.pbs.org/wgbh/nova/photo51/>
- [64] McElheny 2004 p.53: The weekend (January 31-February 1) after seeing photo 51, Watson informed Bragg of the X-ray diffraction image of DNA in B form. Bragg gave them permission to restart their research on DNA (that is, model building).
- [65] McElheny 2004 p.54: On Sunday February 8, 1953, Maurice Wilkes gave Watson and Crick permission to work on models, as Wilkes would not be building models until Franklin left DNA research.
- [66] McElheny 2004 p.56: Jerry Donohue, on sabbatical from Pauling's lab and visiting Cambridge, advises Watson that textbook form of the base pairs was incorrect for DNA base pairs; rather, the keto form of the base pairs should be used instead. This form allowed the bases' hydrogen bonds to pair 'unlike' with 'unlike', rather than to pair 'like' with 'like', as Watson was inclined to model, on the basis of the textbook statements. On February 27, 1953, Watson was convinced enough to make cardboard models of the nucleotides in their keto form.
- [67] "Suddenly I became aware that an adenine-thymine pair held together by two hydrogen bonds was identical in shape to a guanine-cytosine pair held together by at least two hydrogen bonds. ..." —Watson 1968, pp. 194–197.
- McElheny 2004 p.57 Saturday, February 28, 1953, Watson tried 'like with like' and admitted these base pairs didn't have hydrogen bonds that line up. But after trying 'unlike with unlike', and getting Jerry Donohue's approval, the base pairs turned out to be identical in shape (as Watson stated above in his 1968 *Double Helix* memoir quoted above). Watson now felt confident enough to inform Crick. (Of course, 'unlike with unlike' increases the number of possible codons, if this scheme were a genetic code.)
- [68] See, e.g., *Physics Today*, **59**(1), p42. Richmann electrocuted in St. Petersburg (1753) ([http://ptonline.aip.org/journals/doc/PHTOAD-ft/vol\\_59/iss\\_1/42\\_1.shtml?bypassSSO=1](http://ptonline.aip.org/journals/doc/PHTOAD-ft/vol_59/iss_1/42_1.shtml?bypassSSO=1))
- [69] Aristotle, "Prior Analytics", Hugh Tredennick (trans.), pp. 181-531 in *Aristotle, Volume 1*, Loeb Classical Library, William Heinemann, London, UK, 1938.
- [70] Peirce (1877), "The Fixation of Belief", *Popular Science Monthly*, v. 12, pp. 1–15. Reprinted often, including (*Collected Papers of Charles Sanders Peirce* v. 5, paragraphs 358–87), (*The Essential Peirce*, v. 1, pp. 109–23). *Peirce.org* Eprint (<http://www.peirce.org/writings/p107.html>). *Wikisource* Eprint.
- [71] "What one does not in the least doubt one should not pretend to doubt; but a man should train himself to doubt," said Peirce in a brief intellectual autobiography; see Ketner, Kenneth Laine (2009) "Charles Sanders Peirce: Interdisciplinary Scientist" in *The Logic of Interdisciplinarity*). Peirce held that actual, genuine doubt originates externally, usually in surprise, but also that it is to be sought and cultivated, "provided only that it be the weighty and noble metal itself, and no counterfeit nor paper substitute"; in "Issues of Pragmatism", *The Monist*, v. XV, n. 4, pp. 481-99, see p. 484 (<http://www.archive.org/stream/monistquart15hegeuoft#page/484/mode/1up>), and p. 491 (<http://www.archive.org/stream/monistquart15hegeuoft#page/491/mode/1up>). (Reprinted in *Collected Papers* v. 5, paragraphs 438-63, see 443 and 451).
- [72] Peirce (1898), "Philosophy and the Conduct of Life", Lecture 1 of the Cambridge (MA) Conferences Lectures, published in *Collected Papers* v. 1, paragraphs 616-48 in part and in *Reasoning and the Logic of Things*, Ketner (ed., intro.) and Putnam (intro., comm.), pp. 105-22, reprinted in *Essential Peirce* v. 2, pp. 27-41.
- [73] Peirce (1899), "F.R.L." [First Rule of Logic], *Collected Papers* v. 1, paragraphs 135-40, Eprint ([http://www.princeton.edu/~batke/peirce/frl\\_99.htm](http://www.princeton.edu/~batke/peirce/frl_99.htm))
- [74] *Collected Papers* v. 5, in paragraph 582, from 1898:
- ... [rational] inquiry of every type, fully carried out, has the vital power of self-correction and of growth. This is a property so deeply saturating its inmost nature that it may truly be said that there is but one thing needful for learning the truth, and that is a hearty and active desire to learn what is true.
- [75] Peirce (1877), "How to Make Our Ideas Clear", *Popular Science Monthly*, v. 12, pp. 286–302. Reprinted often, including *Collected Papers* v. 5, paragraphs 388–410, *Essential Peirce* v. 1, pp. 124–41. *Arisbe* Eprint (<http://www.cspeirce.com/menu/library/bycsp/ideas/id-frame.htm>). *Wikisource* Eprint.
- [76] Peirce (1868), "Some Consequences of Four Incapacities", *Journal of Speculative Philosophy* v. 2, n. 3, pp. 140–57. Reprinted *Collected Papers* v. 5, paragraphs 264–317, *The Essential Peirce* v. 1, pp. 28–55, and elsewhere. *Arisbe* Eprint (<http://www.cspeirce.com/menu/library/bycsp/conseq/cn-frame.htm>)
- [77] Peirce (1878), "The Doctrine of Chances", *Popular Science Monthly* v. 12, pp. 604-15, see pp. 610 (<http://www.archive.org/stream/popscimonthly12yoummiss#page/618/mode/1up>)-11 via *Internet Archive*. Reprinted *Collected Papers* v. 2, paragraphs 645-68, *Essential Peirce* v. 1, pp. 142-54. "...death makes the number of our risks, the number of our inferences, finite, and so makes their mean result uncertain. The very idea of probability and of reasoning rests on the assumption that this number is indefinitely great. .... ..logicality inexorably requires that our interests shall not be limited. .... Logic is rooted in the social principle."
- [78] Peirce (1908), "A Neglected Argument for the Reality of God", *Hibbert Journal* v. 7, pp. 90-112. s:A Neglected Argument for the Reality of God with added notes. Reprinted with previously unpublished part, *Collected Papers* v. 6, paragraphs 452-85, *The Essential Peirce* v. 2, pp. 434-50, and elsewhere.
- [79] Peirce (c. 1906), "PAP (Prolegomena for an Apology to Pragmatism)" (Manuscript 293, not the like-named article), *The New Elements of Mathematics* (NEM) 4:319-320, see first quote under " Abduction (<http://www.helsinki.fi/science/commens/terms/abduction.html>)" at *Commens Dictionary of Peirce's Terms*.

[80] Peirce, Carnegie application (L75, 1902), *New Elements of Mathematics* v. 4, pp. 37-38:

For it is not sufficient that a hypothesis should be a justifiable one. Any hypothesis which explains the facts is justified critically. But among justifiable hypotheses we have to select that one which is suitable for being tested by experiment.

[81] Peirce (1902), Carnegie application, see MS L75.329-330, from Draft D (<http://www.cspeirce.com/menu/library/bycsp/l75/ver1/175v1-08.htm#m27>) of Memoir 27:

Consequently, to discover is simply to expedite an event that would occur sooner or later, if we had not troubled ourselves to make the discovery. Consequently, the art of discovery is purely a question of economics. The economics of research is, so far as logic is concerned, the leading doctrine with reference to the art of discovery. Consequently, the conduct of abduction, which is chiefly a question of heuritic and is the first question of heuritic, is to be governed by economical considerations.

[82] Peirce (1903), "Pragmatism — The Logic of Abduction", *Collected Papers* v. 5, paragraphs 195-205, especially 196. Eprint (<http://www.textlog.de/7663.html>).

[83] Peirce, "On the Logic of Drawing Ancient History from Documents", *Essential Peirce* v. 2, see pp. 107-9. On Twenty Questions, p. 109:

Thus, twenty skillful hypotheses will ascertain what 200,000 stupid ones might fail to do.

[84] Peirce (1878), "The Probability of Induction", *Popular Science Monthly*, v. 12, pp. 705-18, see 718 (<http://books.google.com/books?id=ZKMVAAAAYAAJ&pg=PA718>) *Google Books*; 718 (<http://www.archive.org/stream/popscimonthly12yoummiss#page/728/mode/1up>) via *Internet Archive*. Reprinted often, including (*Collected Papers* v. 2, paragraphs 669-93), (*The Essential Peirce* v. 1, pp. 155-69).

[85] Peirce (1905 draft "G" of "A Neglected Argument"), "Crude, Quantitative, and Qualitative Induction", *Collected Papers* v. 2, paragraphs 755–760, see 759. Find under "Induction" (<http://www.helsinki.fi/science/commens/terms/induction.html>) at *Commens Dictionary of Peirce's Terms*.

[86] . Brown, C. (2005) Overcoming Barriers to Use of Promising Research Among Elite Middle East Policy Groups, *Journal of Social Behaviour and Personality*, Select Press.

[87] [http://www.nature.com/nature/submit/get\\_published/index.html](http://www.nature.com/nature/submit/get_published/index.html)

[88] Hanson, Norwood (1958), *Patterns of Discovery*, Cambridge University Press, ISBN 0-521-05197-5

[89] Kuhn 1962, p. 113 ISBN 978-1-4432-5544-8

[90] Feyerabend, Paul K (1960) "Patterns of Discovery" *The Philosophical Review* (1960) vol. 69 (2) pp. 247-252

[91] Kuhn, Thomas S., "The Function of Measurement in Modern Physical Science", *ISIS* 52(2), 161–193, 1961.

[92] Feyerabend, Paul K., *Against Method, Outline of an Anarchistic Theory of Knowledge*, 1st published, 1975. Reprinted, Verso, London, UK, 1978.

- *Higher Superstition: The Academic Left and Its Quarrels with Science*, The Johns Hopkins University Press, 1997
- *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science*, Picador; 1st Picador USA Pbk. Ed edition, 1999
- *The Sokal Hoax: The Sham That Shook the Academy*, University of Nebraska Press, 2000 ISBN 0-8032-7995-7
- *A House Built on Sand: Exposing Postmodernist Myths About Science*, Oxford University Press, 2000
- *Intellectual Impostures*, Economist Books, 2003

[94] Dunbar, K., & Fugelsang, J. (2005). Causal thinking in science: How scientists and students interpret the unexpected. In M. E. Gorman, R. D. Tweney, D. Gooding & A. Kincannon (Eds.), *Scientific and Technical Thinking* (pp. 57-79). Mahwah, NJ: Lawrence Erlbaum Associates.

[95] Oliver, J.E. (1991) Ch2. of *The incomplete guide to the art of discovery*. New York:NY, Columbia University Press.

[96] Talib contributes a brief description of anti-fragility, [http://www.edge.org/q2011/q11\\_3.html](http://www.edge.org/q2011/q11_3.html)

[97] Riccardo Pozzo (2004) *The impact of Aristotelianism on modern philosophy* (<http://books.google.com/books?id=vayp8jxcPr0C&pg=&dq&hl=en#v=onepage&q=&f=false>). CUA Press. p.41. ISBN 0-8132-1347-9

[98] The ancient Egyptians observed that heliacal rising of a certain star, *Sothis* (Greek for *Sopdet* (Egyptian), known to the West as *Sirius*), marked the annual flooding of the Nile river. See Neugebauer, Otto (1969) [1957], *The Exact Sciences in Antiquity* (<http://books.google.com/?id=JVhTtVA2zr8C>) (2 ed.), Dover Publications, ISBN 978-0-486-22332-2, , p.82, and also the 1911 *Britannica*, "Egypt".

[99] The Rhind papyrus lists practical examples in arithmetic and geometry —1911 *Britannica*, "Egypt".

[100] The Ebers papyrus lists some of the 'mysteries of the physician', as cited in the 1911 *Britannica*, "Egypt"

[101] Rosanna Gorini (2003), "Al-Haytham the Man of Experience, First Steps in the Science of Vision", *International Society for the History of Islamic Medicine*, Institute of Neurosciences, Laboratory of Psychobiology and Psychopharmacology, Rome, Italy:

"According to the majority of the historians al-Haytham was the pioneer of the modern scientific method. With his book he changed the meaning of the term optics and established experiments as the norm of proof in the field. His investigations are based not on abstract theories, but on experimental evidences and his experiments were systematic and repeatable."

[102] Niccolò Leonicensi (1509), *De Plinii et aliorum erroribus liber* apud Ferrara, as cited by Sanches, Limbrick & Thomson 1988, p. 13

- [103] 'I have sometimes seen a verbose quibbler attempting to persuade some ignorant person that white was black; to which the latter replied, "I do not understand your reasoning, since I have not studied as much as you have; yet I honestly believe that white differs from black. But pray go on refuting me for just as long as you like." '— Sanches, Limbrick & Thomson 1988, p. 276
- [104] Sanches, Limbrick & Thomson 1988, p. 278.
- [105] Bacon, Francis *Novum Organum (The New Organon)*, 1620. Bacon's work described many of the accepted principles, underscoring the importance of empirical results, data gathering and experiment. *Encyclopaedia Britannica* (1911), "Bacon, Francis" states: [In *Novum Organum*, we ] "proceed to apply what is perhaps the most valuable part of the Baconian method, the process of exclusion or rejection. This elimination of the non-essential, ..., is the most important of Bacon's contributions to the logic of induction, and that in which, as he repeatedly says, his method differs from all previous philosophies."
- [106] "John Stuart Mill (Stanford Encyclopedia of Philosophy)" (<http://plato.stanford.edu/entries/mill/#SciMet>). plato.stanford.edu. . Retrieved 2009-07-31.
- [107] Gauch 2003, pp. 52–53
- [108] George Sampson (1970). *The concise Cambridge history of English literature* (<http://books.google.com/books?id=gS3RoUYJi8C&pg=&dq=&hl=en#v=onepage&q=&f=false>). Cambridge University Press. p.174. ISBN 0-521-09581-6
- [109] <http://www.cspeirce.com/menu/library/bycsp/ideas/id-frame.htm>
- [110] *Logik der Forschung*, new appendices \*XVII–\*XIX (not yet available in the English edition *Logic of scientific discovery*)
- [111] *Logic of Scientific discovery*, p. 20
- [112] Karl Popper: On the non-existence of scientific method. *Realism and the Aim of Science* (1983)
- [113] Karl Popper: Science: Conjectures and Refutations. *Conjectures and Refutations*, section VII
- [114] Karl Popper: On knowledge. *In search of a better world*, section II
- [115] "When we are working intensively, we feel keenly the progress of our work; we are elated when our progress is rapid, we are depressed when it is slow." — the mathematician Pólya 1957, p. 131 in the section on 'Modern heuristic'.
- [116] "Philosophy [i.e., physics] is written in this grand book--I mean the universe--which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is written. It is written in the language of mathematics, and its characters are triangles, circles, and other geometrical figures, without which it is humanly impossible to understand a single word of it; without these, one is wandering around in a dark labyrinth." —Galileo Galilei, *Il Saggiatore (The Assayer)*, 1623), as translated by Stillman Drake (1957), *Discoveries and Opinions of Galileo* pp. 237-8, as quoted by di Francia 1981, p. 10.
- [117] Pólya 1957 2nd ed.
- [118] George Pólya (1954), *Mathematics and Plausible Reasoning Volume I: Induction and Analogy in Mathematics*,
- [119] George Pólya (1954), *Mathematics and Plausible Reasoning Volume II: Patterns of Plausible Reasoning*.
- [120] Pólya 1957, p. 142
- [121] Pólya 1957, p. 144
- [122] Mackay 1991 p.100
- [123] See the development, by generations of mathematicians, of Euler's formula for polyhedra as documented by Lakatos, Imre (1976), *Proofs and refutations*, Cambridge: Cambridge University Press, ISBN 0-521-29038-4

## References

- Born, Max (1949), *Natural Philosophy of Cause and Chance*, Peter Smith, also published by Dover, 1964. From the Waynflete Lectures, 1948. On the web. N.B.: the web version does not have the 3 addenda by Born, 1950, 1964, in which he notes that all knowledge is subjective. Born then proposes a solution in Appendix 3 (1964) ([http://www.archive.org/stream/naturalphilosoph032159mbp/naturalphilosoph032159mbp\\_djvu.txt](http://www.archive.org/stream/naturalphilosoph032159mbp/naturalphilosoph032159mbp_djvu.txt))
- Brody, Thomas A. (1993), *The Philosophy Behind Physics*, Springer Verlag, ISBN 0-387-55914-0. (Luis De La Peña and Peter E. Hodgson, eds.)
- Bruno, Leonard C. (1989), *The Landmarks of Science*, ISBN 0-8160-2137-6
- Bynum, W.F.; Porter, Roy (2005), *Oxford Dictionary of Scientific Quotations*, Oxford, ISBN 0-19-858409-1.
- di Francia, G. Toraldo (1981), *The Investigation of the Physical World*, Cambridge University Press, ISBN 0-521-29925-X.
- Einstein, Albert; Infeld, Leopold (1938), *The Evolution of Physics: from early concepts to relativity and quanta*, New York: Simon and Schuster, ISBN 0-671-20156-5
- Feynman, Richard (1965), *The Character of Physical Law*, Cambridge: M.I.T. Press, ISBN 0-262-56003-8.
- Fleck, Ludwik (1979), *Genesis and Development of a Scientific Fact*, Univ. of Chicago, ISBN 0-226-25325-2. (written in German, 1935, *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*) English translation, 1979 ([http://books.google.com/books?id=0KAGUpaUaGYC&printsec=frontcover&dq=Ludwik+Fleck&source=bl&ots=LcJSSRN\\_ym&](http://books.google.com/books?id=0KAGUpaUaGYC&printsec=frontcover&dq=Ludwik+Fleck&source=bl&ots=LcJSSRN_ym&))

- sig=TKrx9GLwFYRGlIprAcDPFnhJIE&hl=en&ei=rbCWTPGpD8Oblgfmw9iiCg&sa=X&oi=book\_result&ct=result&resnum=3&ved=0CB8Q6AEwAg#v=onepage&q&f=false)
- Galileo (1638), *Two New Sciences*, Leiden: Lodewijk Elzevir, ISBN 0-486-60099-8 Translated from Italian to English in 1914 by Henry Crew and Alfonso de Salvio. Introduction by Antonio Favaro. xxv+300 pages, index. New York: Macmillan, with later reprintings by Dover.
  - Gauch, Hugh G., Jr. (2003), *Scientific Method in Practice* (<http://books.google.com/?id=iVkguqNG9dAC>), Cambridge University Press, ISBN 0-521-01708-4 435 pages
  - Glen, William (ed.) (1994), *The Mass-Extinction Debates: How Science Works in a Crisis*, Stanford, CA: Stanford University Press, ISBN 0-8047-2285-4.
  - Godfrey-Smith, Peter (2003), *Theory and Reality: An introduction to the philosophy of science*, University of Chicago Press, ISBN 0-226-30063-3.
  - Goldhaber, Alfred Scharff; Nieto, Michael Martin (January–March 2010), "Photon and graviton mass limits", *Rev. Mod. Phys.* (American Physical Society) **82**: 939, doi:10.1103/RevModPhys.82.939. pages 939-979.
  - Jevons, William Stanley (1874), *The Principles of Science: A Treatise on Logic and Scientific Method*, Dover Publications, ISBN 1-4304-8775-5. 1877, 1879. Reprinted with a foreword by Ernst Nagel, New York, NY, 1958.
  - Kuhn, Thomas S. (1962), *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press. 2nd edition 1970. 3rd edition 1996.
  - Mackay, Alan L. (ed.) (1991), *Dictionary of Scientific Quotations*, London: IOP Publishing Ltd, ISBN 0-7503-0106-6
  - McElheny, Victor K. (2004), *Watson & DNA: Making a scientific revolution*, Basic Books, ISBN 0-7382-0866-3.
  - Moulton, Forest Ray; Schifferes, Justus J. (eds., Second Edition) (1960), *The Autobiography of Science*, Doubleday.
  - Needham, Joseph; Wang, Ling (王玲) (1954), *Science and Civilisation in China, 1 Introductory Orientations*, Cambridge University Press
  - Newton, Isaac (1687, 1713, 1726), *Philosophiæ Naturalis Principia Mathematica*, University of California Press, ISBN 0-520-08817-4, Third edition. From I. Bernard Cohen and Anne Whitman's 1999 translation, 974 pages.
  - Ørsted, Hans Christian (1997), *Selected Scientific Works of Hans Christian Ørsted*, Princeton, ISBN 0-691-04334-5. Translated to English by Karen Jelved, Andrew D. Jackson, and Ole Knudsen, (translators 1997).
  - Peirce, C. S. — see Charles Sanders Peirce bibliography.
  - Poincaré, Henri (1905), *Science and Hypothesis* Eprint ([http://www.brocku.ca/MeadProject/Poincare/Poincare\\_1905\\_toc.html](http://www.brocku.ca/MeadProject/Poincare/Poincare_1905_toc.html))
  - Pólya, George (1957), *How to Solve It*, Princeton University Press, ISBN -691-08097-6
  - Popper, Karl R., *The Logic of Scientific Discovery*, 1934, 1959.
  - Sambursky, Shmuel (ed.) (1974), *Physical Thought from the Presocratics to the Quantum Physicists*, Pica Press, ISBN 0-87663-712-8.
  - Sanches, Francisco; Limbrick, Elaine. Introduction, Notes, and Bibliography; Thomson, Douglas F.S. Latin text established, annotated, and translated. (1988), *That Nothing is Known*, Cambridge: Cambridge University Press, ISBN 0-521-35077-8 Critical edition.
  - Taleb, Nassim Nicholas (2007), *The Black Swan*, Random House, ISBN 978-1-4000-6351-2
  - Watson, James D. (1968), *The Double Helix*, New York: Atheneum, Library of Congress card number 68-16217.
-

## Further reading

- Bauer, Henry H., *Scientific Literacy and the Myth of the Scientific Method*, University of Illinois Press, Champaign, IL, 1992
- Beveridge, William I. B., *The Art of Scientific Investigation*, Heinemann, Melbourne, Australia, 1950.
- Bernstein, Richard J., *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*, University of Pennsylvania Press, Philadelphia, PA, 1983.
- Brody, Baruch A. and Capaldi, Nicholas, *Science: Men, Methods, Goals: A Reader: Methods of Physical Science* (<http://books.google.com/books?id=d1heAAAIAAJ&pgis=1>), W. A. Benjamin, 1968
- Brody, Baruch A., and Grandy, Richard E., *Readings in the Philosophy of Science*, 2nd edition, Prentice Hall, Englewood Cliffs, NJ, 1989.
- Burks, Arthur W., *Chance, Cause, Reason — An Inquiry into the Nature of Scientific Evidence*, University of Chicago Press, Chicago, IL, 1977.
- Alan Chalmers. *What is this thing called science?*. Queensland University Press and Open University Press, 1976.
- Crick, Francis (1988), *What Mad Pursuit: A Personal View of Scientific Discovery*, New York: Basic Books, ISBN 0-465-09137-7.
- Dewey, John, *How We Think*, D.C. Heath, Lexington, MA, 1910. Reprinted, Prometheus Books, Buffalo, NY, 1991.
- Earman, John (ed.), *Inference, Explanation, and Other Frustrations: Essays in the Philosophy of Science*, University of California Press, Berkeley & Los Angeles, CA, 1992.
- Fraassen, Bas C. van, *The Scientific Image*, Oxford University Press, Oxford, UK, 1980.
- Franklin, James (2009), *What Science Knows: And How It Knows It*, New York: Encounter Books, ISBN 1-59403-207-6.
- Gadamer, Hans-Georg, *Reason in the Age of Science*, Frederick G. Lawrence (trans.), MIT Press, Cambridge, MA, 1981.
- Giere, Ronald N. (ed.), *Cognitive Models of Science*, vol. 15 in 'Minnesota Studies in the Philosophy of Science', University of Minnesota Press, Minneapolis, MN, 1992.
- Hacking, Ian, *Representing and Intervening, Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press, Cambridge, UK, 1983.
- Heisenberg, Werner, *Physics and Beyond, Encounters and Conversations*, A.J. Pomerans (trans.), Harper and Row, New York, NY 1971, pp. 63–64.
- Holton, Gerald, *Thematic Origins of Scientific Thought, Kepler to Einstein*, 1st edition 1973, revised edition, Harvard University Press, Cambridge, MA, 1988.
- Kuhn, Thomas S., *The Essential Tension, Selected Studies in Scientific Tradition and Change*, University of Chicago Press, Chicago, IL, 1977.
- Latour, Bruno, *Science in Action, How to Follow Scientists and Engineers through Society*, Harvard University Press, Cambridge, MA, 1987.
- Losee, John, *A Historical Introduction to the Philosophy of Science*, Oxford University Press, Oxford, UK, 1972. 2nd edition, 1980.
- Maxwell, Nicholas, *The Comprehensibility of the Universe: A New Conception of Science*, Oxford University Press, Oxford, 1998. Paperback 2003.
- McCarty, Maclyn (1985), *The Transforming Principle: Discovering that genes are made of DNA*, New York: W. W. Norton, pp. 252 , ISBN 0-393-30450-7. Memoir of a researcher in the Avery–MacLeod–McCarty experiment.
- McComas, William F., ed. *The Principal Elements of the Nature of Science: Dispelling the Myths* (<http://coehp.uark.edu/pase/TheMythsOfScience.pdf>) PDF (189 KB), from *The Nature of Science in Science Education*, pp53–70, Kluwer Academic Publishers, Netherlands 1998.

- Misak, Cheryl J., *Truth and the End of Inquiry, A Peircean Account of Truth*, Oxford University Press, Oxford, UK, 1991.
- Piattelli-Palmarini, Massimo (ed.), *Language and Learning, The Debate between Jean Piaget and Noam Chomsky*, Harvard University Press, Cambridge, MA, 1980.
- Popper, Karl R., *Unended Quest, An Intellectual Autobiography*, Open Court, La Salle, IL, 1982.
- Putnam, Hilary, *Renewing Philosophy*, Harvard University Press, Cambridge, MA, 1992.
- Rorty, Richard, *Philosophy and the Mirror of Nature*, Princeton University Press, Princeton, NJ, 1979.
- Salmon, Wesley C., *Four Decades of Scientific Explanation*, University of Minnesota Press, Minneapolis, MN, 1990.
- Shimony, Abner, *Search for a Naturalistic World View: Vol. 1, Scientific Method and Epistemology, Vol. 2, Natural Science and Metaphysics*, Cambridge University Press, Cambridge, UK, 1993.
- Thagard, Paul, *Conceptual Revolutions*, Princeton University Press, Princeton, NJ, 1992.
- Ziman, John (2000). *Real Science: what it is, and what it means*. Cambridge, UK: Cambridge University Press.

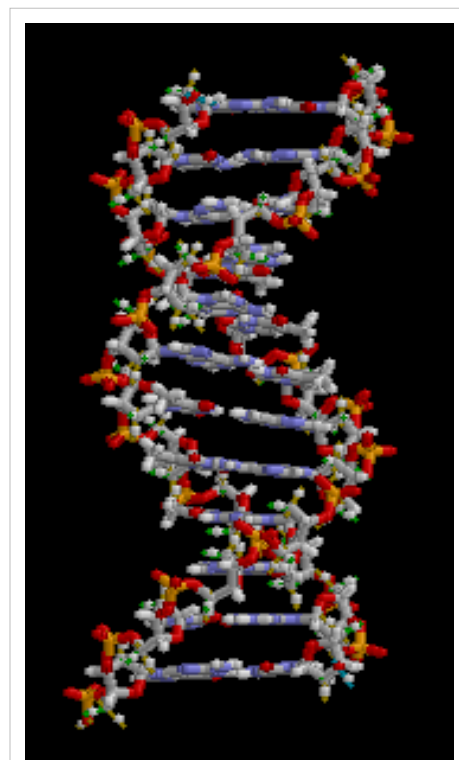
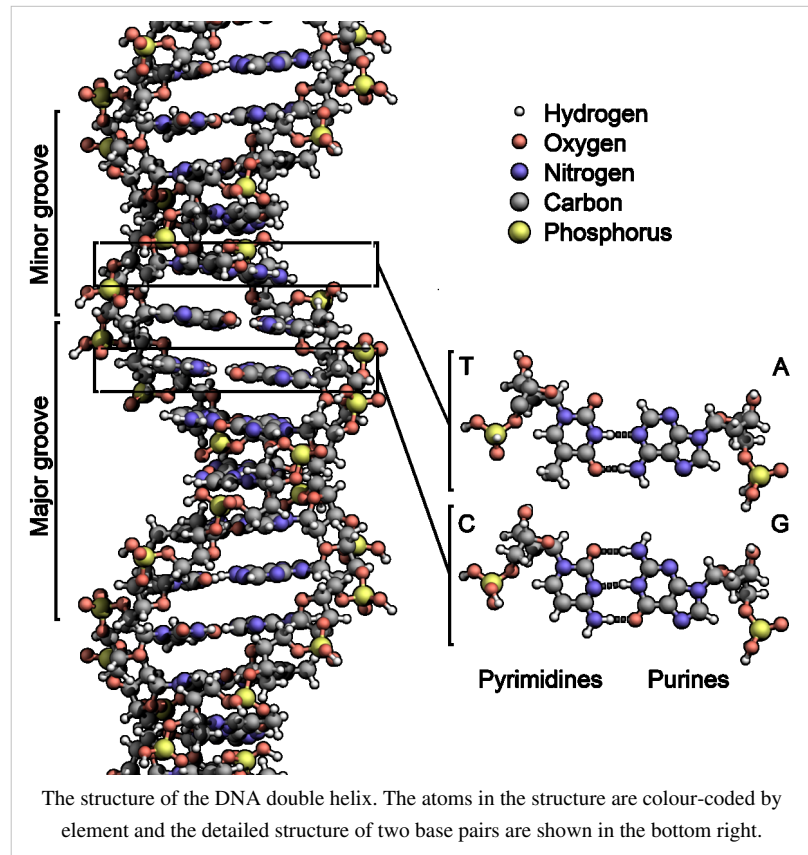
## External links

- Scientific method (<http://philpapers.org/browse/scientific-method>) at PhilPapers
- Scientific method (<https://inpho.cogs.indiana.edu/idea/1916>) at the Indiana Philosophy Ontology Project
- An Introduction to Science: Scientific Thinking and a scientific method (<http://www.geo.sunysb.edu/esp/files/scientific-method.html>) by Steven D. Schafersman.
- Introduction to the scientific method ([http://teacher.nsr.l.rochester.edu/phy\\_labs/AppendixE/AppendixE.html](http://teacher.nsr.l.rochester.edu/phy_labs/AppendixE/AppendixE.html)) at the University of Rochester
- Theory-ladenness (<http://www.galilean-library.org/theory.html>) by Paul Newall at The Galilean Library
- Lecture on Scientific Method by Greg Anderson (<http://web.archive.org/web/20060428080832/http://pasadena.wr.usgs.gov/office/ganderson/es10/lectures/lecture01/lecture01.html>)
- Using the scientific method for designing science fair projects ([http://www.sciencemadesimple.com/scientific\\_method.html](http://www.sciencemadesimple.com/scientific_method.html))
- SCIENTIFIC METHODS an online book by Richard D. Jarrard (<http://emotionalcompetency.com/sci/booktoc.html>)
- Richard Feynman on the Key to Science (<http://www.youtube.com/watch?v=b240PGCMwV0>) (one minute, three seconds), from the Cornell Lectures.
- Lectures on the Scientific Method (<http://www.dbskeptic.com/2010/03/14/what-it-means-to-be-scientifically-proven/>) by Nick Josh Karean, Kevin Padian, Michael Shermer and Richard Dawkins

# DNA

**Deoxyribonucleic acid (DNA)** is a molecule encoding the genetic instructions used in the development and functioning of all known living organisms and many viruses. Along with RNA and proteins, DNA is one of the three major macromolecules that are essential for all known forms of life. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine) recorded using the letters G, A, T, and C. Most DNA molecules are double-stranded helices, consisting of two long polymers of simple units called nucleotides, molecules with backbones made of alternating sugars (deoxyribose) and phosphate groups (related to phosphoric acid), with the nucleobases (G, A, T, C) attached to the sugars. DNA is well-suited for biological information storage, since the DNA backbone is resistant to cleavage and the double-stranded structure provides the molecule with a built-in duplicate of the encoded information.

These two strands run in opposite directions to each other and are therefore anti-parallel, one backbone being 3' (three prime) and the other 5' (five prime). This refers to the direction the 3rd and 5th carbon on the sugar molecule is facing. Attached to each sugar is one of four types of molecules called nucleobases (informally, *bases*). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within



The structure of part of a DNA double helix



proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription.

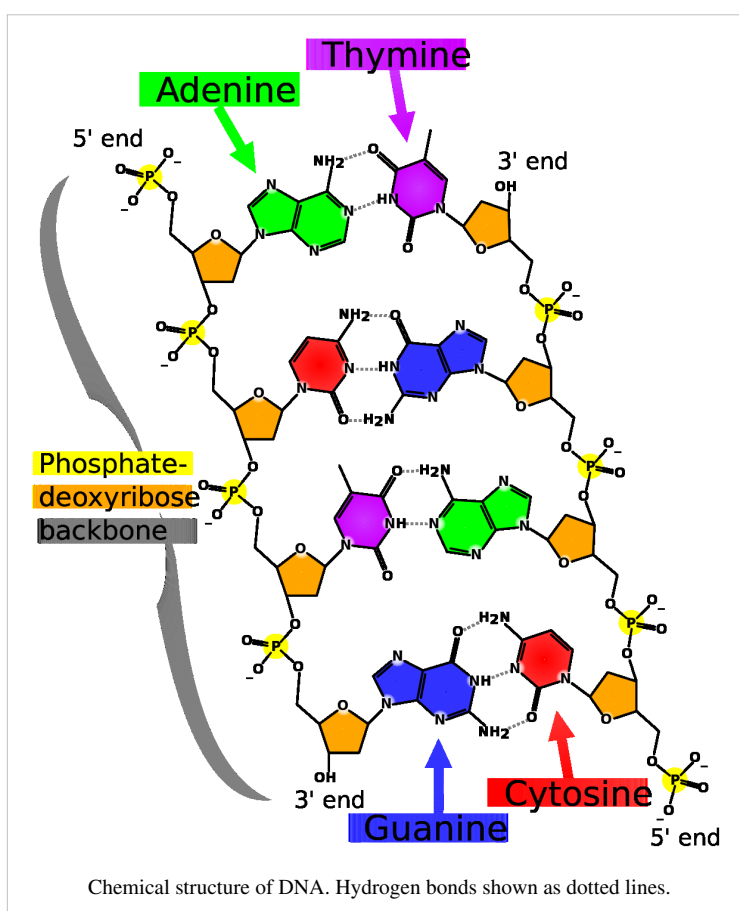
Within cells, DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts.<sup>[1]</sup> In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

## Properties

DNA is a long polymer made from repeating units called nucleotides.<sup>[2][3][4]</sup>

DNA was first identified and isolated by Friedrich Miescher and the double helix structure of DNA was first discovered by James D. Watson and Francis Crick. The structure of DNA of all species comprises two helical chains each coiled round the same axis, and each with a pitch of 34 ångströms (3.4 nanometres) and a radius of 10 ångströms (1.0 nanometres).<sup>[5]</sup>

According to another study, when measured in a particular solution, the DNA chain measured 22 to 26 ångströms wide (2.2 to 2.6 nanometres), and one nucleotide unit measured 3.3 Å (0.33 nm) long.<sup>[6]</sup> Although each individual repeating unit is very small, DNA polymers can be very large molecules containing millions of nucleotides. For instance, the largest human chromosome, chromosome number 1, is approximately 220 million base pairs long.<sup>[7]</sup>



In living organisms DNA does not usually exist as a single molecule, but instead as a pair of molecules that are held tightly together.<sup>[8][9]</sup> These two long strands entwine like vines, in the shape of a double helix. The nucleotide repeats contain both the segment of the backbone of the molecule, which holds the chain together, and a nucleobase, which interacts with the other DNA strand in the helix. A nucleobase linked to a sugar is called a nucleoside and a base linked to a sugar and one or more phosphate groups is called a nucleotide. A polymer comprising multiple linked nucleotides (as in DNA) is called a polynucleotide.<sup>[10]</sup>

The backbone of the DNA strand is made from alternating phosphate and sugar residues.<sup>[11]</sup> The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*. The asymmetric ends of DNA strands are called the 5'

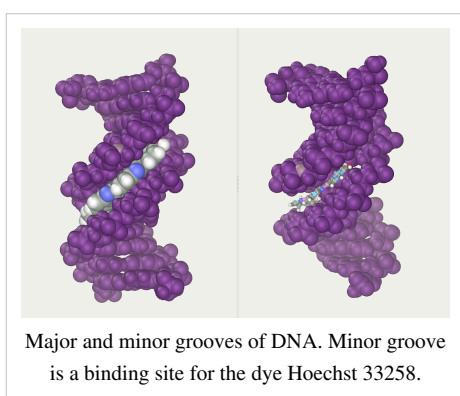
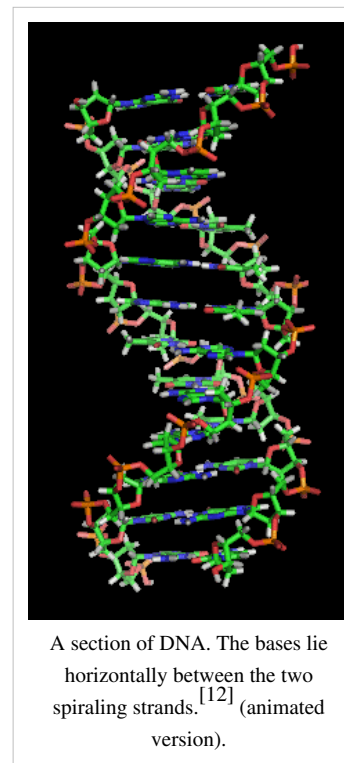
(*five prime*) and 3' (*three prime*) ends, with the 5' end having a terminal phosphate group and the 3' end a terminal hydroxyl group. One major difference between DNA and RNA is the sugar, with the 2-deoxyribose in DNA being replaced by the alternative pentose sugar ribose in RNA.<sup>[9]</sup>

The DNA double helix is stabilized primarily by two forces: hydrogen bonds between nucleotides and base-stacking interactions among aromatic nucleobases.<sup>[13]</sup> In the aqueous environment of the cell, the conjugated  $\pi$  bonds of nucleotide bases align perpendicular to the axis of the DNA molecule, minimizing their interaction with the solvation shell and therefore, the Gibbs free energy. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide, as shown for adenosine monophosphate.

### Nucleobase classification

The nucleobases are classified into two types: the purines, A and G, being fused five- and six-membered heterocyclic compounds, and the pyrimidines, the six-membered rings C and T.<sup>[9]</sup> A fifth pyrimidine nucleobase, uracil (U), usually takes the place of thymine in RNA and differs from thymine by lacking a methyl group on its ring. In addition to RNA and DNA a large number of artificial nucleic acid analogues have also been created to study the properties of nucleic acids, or for use in biotechnology.<sup>[14]</sup>

Uracil is not usually found in DNA, occurring only as a breakdown product of cytosine. However in a number of bacteriophages – *Bacillus subtilis* bacteriophages PBS1 and PBS2 and *Yersinia* bacteriophage piR1-37 – thymine has been replaced by uracil.<sup>[15]</sup> A modified form (beta-d-glucopyranosyloxymethyluracil) is also found in a number of organisms: the flagellates *Diplonema* and *Euglena*, and all the kinetoplastid genera<sup>[16]</sup> Biosynthesis of J occurs in two steps: in the first step a specific thymidine in DNA is converted into hydroxymethyldeoxyuridine; in the second HOMedU is glycosylated to form J.<sup>[17]</sup> Proteins that bind specifically to this base have been identified.<sup>[18][19][20]</sup> These proteins appear to be distant relatives of the Tet1 oncogene that is involved in the pathogenesis of acute myeloid leukemia.<sup>[21]</sup> J appears to act as a termination signal for RNA polymerase II.<sup>[22][23]</sup>

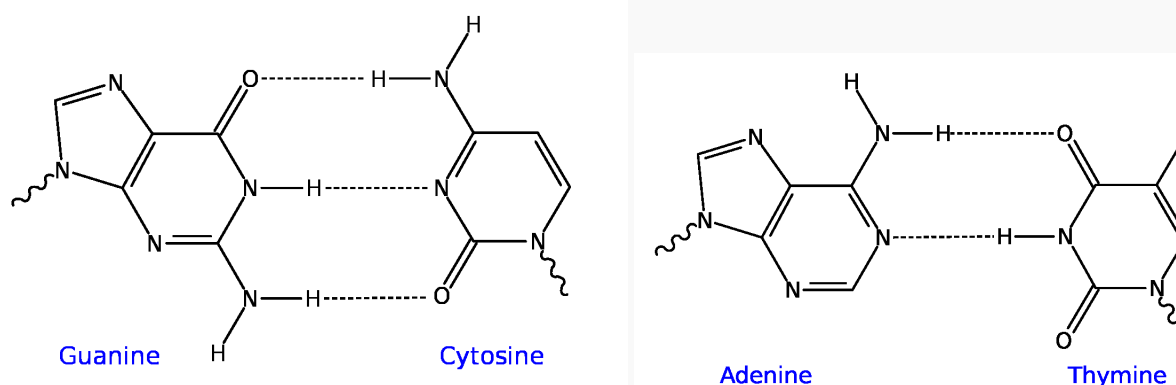


### Grooves

Twin helical strands form the DNA backbone. Another double helix may be found tracing the spaces, or grooves, between the strands. These voids are adjacent to the base pairs and may provide a binding site. As the strands are not symmetrically located with respect to each other, the grooves are unequally sized. One groove, the major groove, is 22 Å wide and the other, the minor groove, is 12 Å wide.<sup>[24]</sup> The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result, proteins like transcription factors that can bind to specific sequences in double-stranded DNA usually make contacts to the sides of the bases exposed in the major groove.<sup>[25]</sup> This situation varies in unusual conformations of DNA within the cell (*see below*), but the major and minor grooves are always named to reflect the differences in size that would be seen if the DNA is twisted back into the ordinary B form.

## Base pairing

In a DNA double helix, each type of nucleobase on one strand bonds with just one type of nucleobase on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with adenine bonding only to thymine in two hydrogen bonds, and cytosine bonding only to guanine in three hydrogen bonds. This arrangement of two nucleotides binding together across the double helix is called a base pair. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature.<sup>[26]</sup> As a result of this complementarity, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication. Indeed, this reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in living organisms.<sup>[3]</sup>



Top, a GC base pair with three hydrogen bonds. Bottom, an AT base pair with two hydrogen bonds. Non-covalent hydrogen bonds between the pairs are shown as dashed lines.

The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds (see figures, right). DNA with high GC-content is more stable than DNA with low GC-content.

As noted above, most DNA molecules are actually two polymer strands, bound together in a helical fashion by noncovalent bonds; this double stranded structure (dsDNA) is maintained largely by the intrastrand base stacking interactions, which are strongest for G,C stacks. The two strands can come apart – a process known as melting – to form two ssDNA molecules. Melting occurs when conditions favor ssDNA; such conditions are high temperature, low salt and high pH (low pH also melts DNA, but since DNA is unstable due to acid depurination, low pH is rarely used).

The stability of the dsDNA form depends not only on the GC-content (% G,C basepairs) but also on sequence (since stacking is sequence specific) and also length (longer molecules are more stable). The stability can be measured in various ways; a common way is the "melting temperature", which is the temperature at which 50% of the ds molecules are converted to ss molecules; melting temperature is dependent on ionic strength and the concentration of DNA. As a result, it is both the percentage of GC base pairs and the overall length of a DNA double helix that determines the strength of the association between the two strands of DNA. Long DNA helices with a high GC-content have stronger-interacting strands, while short helices with high AT content have weaker-interacting strands.<sup>[27]</sup> In biology, parts of the DNA double helix that need to separate easily, such as the TATAAT Pribnow box in some promoters, tend to have a high AT content, making the strands easier to pull apart.<sup>[28]</sup>

In the laboratory, the strength of this interaction can be measured by finding the temperature necessary to break the hydrogen bonds, their melting temperature (also called  $T_m$  value). When all the base pairs in a DNA double helix melt, the strands separate and exist in solution as two entirely independent molecules. These single-stranded DNA molecules (ssDNA) have no single common shape, but some conformations are more stable than others.<sup>[29]</sup>

## Sense and antisense

A DNA sequence is called "sense" if its sequence is the same as that of a messenger RNA copy that is translated into protein.<sup>[30]</sup> The sequence on the opposite strand is called the "antisense" sequence. Both sense and antisense sequences can exist on different parts of the same strand of DNA (i.e. both strands contain both sense and antisense sequences). In both prokaryotes and eukaryotes, antisense RNA sequences are produced, but the functions of these RNAs are not entirely clear.<sup>[31]</sup> One proposal is that antisense RNAs are involved in regulating gene expression through RNA-RNA base pairing.<sup>[32]</sup>

A few DNA sequences in prokaryotes and eukaryotes, and more in plasmids and viruses, blur the distinction between sense and antisense strands by having overlapping genes.<sup>[33]</sup> In these cases, some DNA sequences do double duty, encoding one protein when read along one strand, and a second protein when read in the opposite direction along the other strand. In bacteria, this overlap may be involved in the regulation of gene transcription,<sup>[34]</sup> while in viruses, overlapping genes increase the amount of information that can be encoded within the small viral genome.<sup>[35]</sup>

## Supercoiling

DNA can be twisted like a rope in a process called DNA supercoiling. With DNA in its "relaxed" state, a strand usually circles the axis of the double helix once every 10.4 base pairs, but if the DNA is twisted the strands become more tightly or more loosely wound.<sup>[36]</sup> If the DNA is twisted in the direction of the helix, this is positive supercoiling, and the bases are held more tightly together. If they are twisted in the opposite direction, this is negative supercoiling, and the bases come apart more easily. In nature, most DNA has slight negative supercoiling that is introduced by enzymes called topoisomerases.<sup>[37]</sup> These enzymes are also needed to relieve the twisting stresses introduced into DNA strands during processes such as transcription and DNA replication.<sup>[38]</sup>

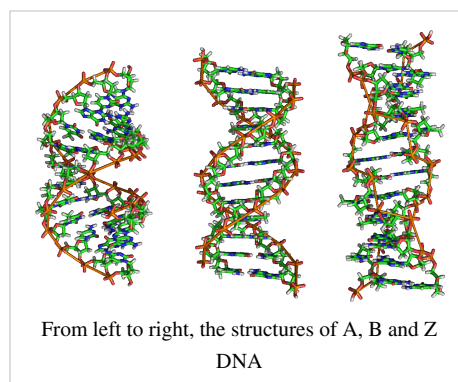
## Alternate DNA structures

DNA exists in many possible conformations that include A-DNA, B-DNA, and Z-DNA forms, although, only B-DNA and Z-DNA have been directly observed in functional organisms.<sup>[11]</sup> The conformation that DNA adopts depends on the hydration level, DNA sequence, the amount and direction of supercoiling, chemical modifications of the bases, the type and concentration of metal ions, as well as the presence of polyamines in solution.<sup>[39]</sup>

The first published reports of A-DNA X-ray diffraction patterns— and also B-DNA — used analyses based on Patterson transforms that provided only a limited amount of structural information for oriented fibers of DNA.<sup>[40][41]</sup> An alternate analysis was then proposed by Wilkins *et al.*, in 1953, for the *in vivo* B-DNA X-ray diffraction/scattering patterns of highly hydrated DNA fibers in terms of squares of Bessel functions.<sup>[42]</sup> In the same journal, James D. Watson and Francis Crick presented their molecular modeling analysis of the DNA X-ray diffraction patterns to suggest that the structure was a double-helix.<sup>[5]</sup>

Although the 'B-DNA form' is most common under the conditions found in cells,<sup>[43]</sup> it is not a well-defined conformation but a family of related DNA conformations<sup>[44]</sup> that occur at the high hydration levels present in living cells. Their corresponding X-ray diffraction and scattering patterns are characteristic of molecular paracrystals with a significant degree of disorder.<sup>[45][46]</sup>

Compared to B-DNA, the A-DNA form is a wider right-handed spiral, with a shallow, wide minor groove and a narrower, deeper major groove. The A form occurs under non-physiological conditions in partially dehydrated samples of DNA, while in the cell it may be produced in hybrid pairings of DNA and RNA strands, as well as in enzyme-DNA complexes.<sup>[47][48]</sup> Segments of DNA where the bases have been chemically modified by methylation



may undergo a larger change in conformation and adopt the Z form. Here, the strands turn about the helical axis in a left-handed spiral, the opposite of the more common B form.<sup>[49]</sup> These unusual structures can be recognized by specific Z-DNA binding proteins and may be involved in the regulation of transcription.<sup>[50]</sup>

### Alternate DNA chemistry

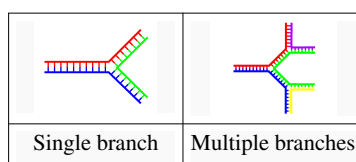
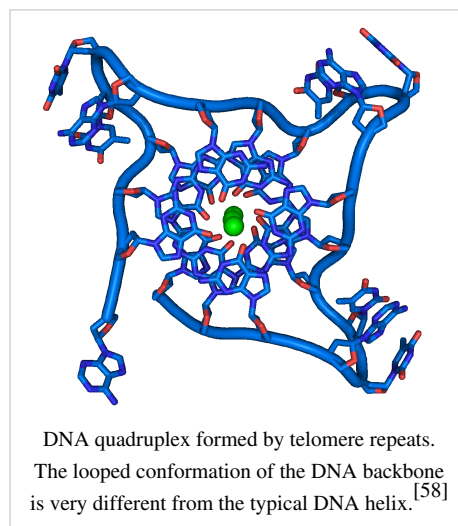
For a number of years exobiologists have proposed the existence of a shadow biosphere, a postulated microbial biosphere of Earth that uses radically different biochemical and molecular processes than currently known life. One of the proposals was the existence of lifeforms that use arsenic instead of phosphorus in DNA. A report in 2010 of the possibility in the bacterium GFAJ-1, was announced,<sup>[51][51][52]</sup> though the research was disputed,<sup>[52][53]</sup> and evidence suggests the bacterium actively prevents the incorporation of arsenic into the DNA backbone and other biomolecules.<sup>[54]</sup>

### Quadruplex structures

At the ends of the linear chromosomes are specialized regions of DNA called telomeres. The main function of these regions is to allow the cell to replicate chromosome ends using the enzyme telomerase, as the enzymes that normally replicate DNA cannot copy the extreme 3' ends of chromosomes.<sup>[55]</sup> These specialized chromosome caps also help protect the DNA ends, and stop the DNA repair systems in the cell from treating them as damage to be corrected.<sup>[56]</sup> In human cells, telomeres are usually lengths of single-stranded DNA containing several thousand repeats of a simple TTAGGG sequence.<sup>[57]</sup>

These guanine-rich sequences may stabilize chromosome ends by forming structures of stacked sets of four-base units, rather than the usual base pairs found in other DNA molecules. Here, four guanine bases form a flat plate and these flat four-base units then stack on top of each other, to form a stable G-quadruplex structure.<sup>[59]</sup> These structures are stabilized by hydrogen bonding between the edges of the bases and chelation of a metal ion in the centre of each four-base unit.<sup>[60]</sup> Other structures can also be formed, with the central set of four bases coming from either a single strand folded around the bases, or several different parallel strands, each contributing one base to the central structure.

In addition to these stacked structures, telomeres also form large loop structures called telomere loops, or T-loops. Here, the single-stranded DNA curls around in a long circle stabilized by telomere-binding proteins.<sup>[61]</sup> At the very end of the T-loop, the single-stranded telomere DNA is held onto a region of double-stranded DNA by the telomere strand disrupting the double-helical DNA and base pairing to one of the two strands. This triple-stranded structure is called a displacement loop or D-loop.<sup>[59]</sup>



Branched DNA can form networks containing multiple branches.

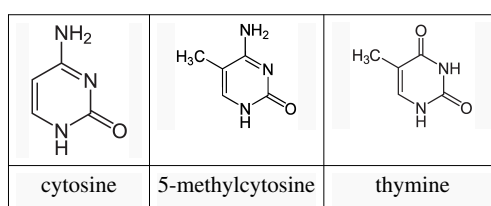
## Branched DNA

In DNA fraying occurs when non-complementary regions exist at the end of an otherwise complementary double-strand of DNA. However, branched DNA can occur if a third strand of DNA is introduced and contains adjoining regions able to hybridize with the frayed regions of the pre-existing double-strand. Although the simplest example of branched DNA involves only three strands of DNA, complexes involving additional strands and multiple branches are also possible.<sup>[62]</sup> Branched DNA can be used in nanotechnology to construct geometric shapes, see the section on uses in technology below.

## Vibration

DNA may carry out low-frequency collective motion as observed by the Raman spectroscopy<sup>[63][64]</sup> and analyzed with a quasi-continuum model.<sup>[65][66]</sup>

## Chemical modifications and altered DNA packaging



Structure of cytosine with and without the 5-methyl group. Deamination converts 5-methylcytosine into thymine.

## Base modifications and DNA packaging

The expression of genes is influenced by how the DNA is packaged in chromosomes, in a structure called chromatin. Base modifications can be involved in packaging, with regions that have low or no gene expression usually containing high levels of methylation of cytosine bases. DNA packaging and its influence on gene expression can also occur by covalent modifications of the histone protein core around which DNA is wrapped in the chromatin structure or else by remodeling carried out by chromatin remodeling complexes (see Chromatin remodeling). There is, further, crosstalk between DNA methylation and histone modification, so they can coordinately affect chromatin and gene expression.<sup>[67]</sup>

For one example, cytosine methylation, produces 5-methylcytosine, which is important for X-chromosome inactivation.<sup>[68]</sup> The average level of methylation varies between organisms – the worm *Caenorhabditis elegans* lacks cytosine methylation, while vertebrates have higher levels, with up to 1% of their DNA containing 5-methylcytosine.<sup>[69]</sup> Despite the importance of 5-methylcytosine, it can deaminate to leave a thymine base, so methylated cytosines are particularly prone to mutations.<sup>[70]</sup> Other base modifications include adenine methylation in bacteria, the presence of 5-hydroxymethylcytosine in the brain,<sup>[71]</sup> and the glycosylation of uracil to produce the "J-base" in kinetoplastids.<sup>[72][73]</sup>



## Damage

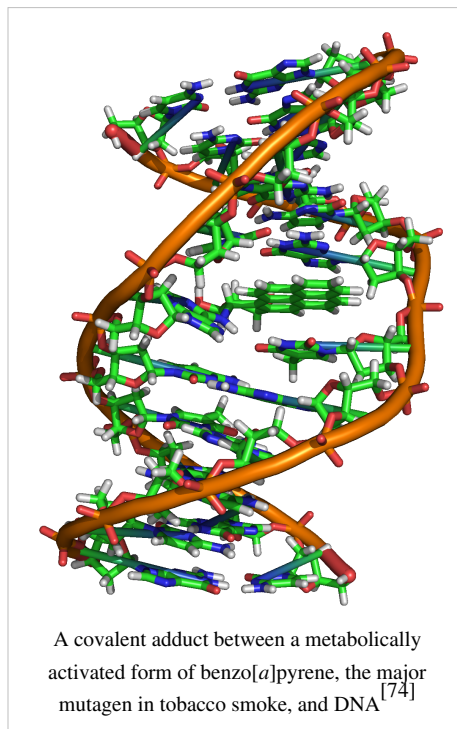
DNA can be damaged by many sorts of mutagens, which change the DNA sequence. Mutagens include oxidizing agents, alkylating agents and also high-energy electromagnetic radiation such as ultraviolet light and X-rays. The type of DNA damage produced depends on the type of mutagen. For example, UV light can damage DNA by producing thymine dimers, which are cross-links between pyrimidine bases.<sup>[75]</sup> On the other hand, oxidants such as free radicals or hydrogen peroxide produce multiple forms of damage, including base modifications, particularly of guanosine, and double-strand breaks.<sup>[76]</sup> A typical human cell contains about 150,000 bases that have suffered oxidative damage.<sup>[77]</sup> Of these oxidative lesions, the most dangerous are double-strand breaks, as these are difficult to repair and can produce point mutations, insertions and deletions from the DNA sequence, as well as chromosomal translocations.<sup>[78]</sup> These mutations can cause cancer. Because of inherent limitations in the DNA repair mechanisms, if humans lived long enough, they would all eventually develop cancer.<sup>[79][80]</sup> DNA damages that are naturally occurring, due to normal cellular processes that produce reactive oxygen species, the hydrolytic activities of cellular water, etc., also occur frequently.

Although most of these damages are repaired, in any cell some DNA damage may remain despite the action of repair processes. These remaining DNA damages accumulate with age in mammalian postmitotic tissues. This accumulation appears to be an important underlying cause of aging.<sup>[81][82][83]</sup>

Many mutagens fit into the space between two adjacent base pairs, this is called *intercalation*. Most intercalators are aromatic and planar molecules; examples include ethidium bromide, acridines, daunomycin, and doxorubicin. In order for an intercalator to fit between base pairs, the bases must separate, distorting the DNA strands by unwinding of the double helix. This inhibits both transcription and DNA replication, causing toxicity and mutations.<sup>[84]</sup> As a result, DNA intercalators may be carcinogens, and in the case of thalidomide, a teratogen.<sup>[85]</sup> Others such as benzo[*a*]pyrene diol epoxide and aflatoxin form DNA adducts which induce errors in replication.<sup>[86]</sup> Nevertheless, due to their ability to inhibit DNA transcription and replication, other similar toxins are also used in chemotherapy to inhibit rapidly growing cancer cells.<sup>[87]</sup>

## Biological functions

DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. The set of chromosomes in a cell makes up its genome; the human genome has approximately 3 billion base pairs of DNA arranged into 46 chromosomes.<sup>[88]</sup> The information carried by DNA is held in the sequence of pieces of DNA called genes. Transmission of genetic information in genes is achieved via complementary base pairing. For example, in transcription, when a cell uses the information in a gene, the DNA sequence is copied into a complementary RNA sequence through the attraction between the DNA and the correct RNA nucleotides. Usually, this RNA copy is then used to make a matching protein sequence in a process called translation, which depends on the same interaction between RNA nucleotides. In alternative fashion, a cell may simply copy its genetic information in a process called DNA replication. The details of these functions are covered in other articles; here we focus on the interactions between DNA and other molecules that mediate the function of the genome.

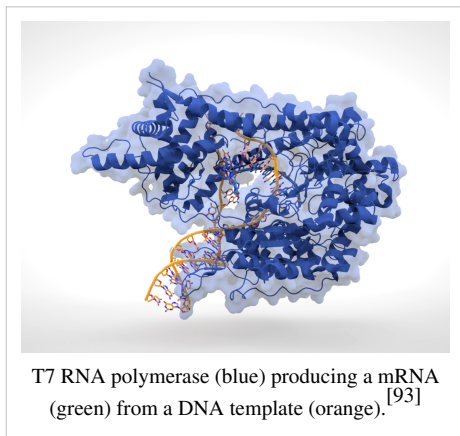




## Genes and genomes

Genomic DNA is tightly and orderly packed in the process called DNA condensation to fit the small available volumes of the cell. In eukaryotes, DNA is located in the cell nucleus, as well as small amounts in mitochondria and chloroplasts. In prokaryotes, the DNA is held within an irregularly shaped body in the cytoplasm called the nucleoid.<sup>[89]</sup> The genetic information in a genome is held within genes, and the complete set of this information in an organism is called its genotype. A gene is a unit of heredity and is a region of DNA that influences a particular characteristic in an organism. Genes contain an open reading frame that can be transcribed, as well as regulatory sequences such as promoters and enhancers, which control the transcription of the open reading frame.

In many species, only a small fraction of the total sequence of the genome encodes protein. For example, only about 1.5% of the human genome consists of protein-coding exons, with over 50% of human DNA consisting of non-coding repetitive sequences.<sup>[90]</sup> The reasons for the presence of so much noncoding DNA in eukaryotic genomes and the extraordinary differences in genome size, or *C-value*, among species represent a long-standing puzzle known as the "C-value enigma".<sup>[91]</sup> However, some DNA sequences that do not code protein may still encode functional non-coding RNA molecules, which are involved in the regulation of gene expression.<sup>[92]</sup>



Some noncoding DNA sequences play structural roles in chromosomes. Telomeres and centromeres typically contain few genes, but are important for the function and stability of chromosomes.<sup>[56][94]</sup> An abundant form of noncoding DNA in humans are pseudogenes, which are copies of genes that have been disabled by mutation.<sup>[95]</sup> These sequences are usually just molecular fossils, although they can occasionally serve as raw genetic material for the creation of new genes through the process of gene duplication and divergence.<sup>[96]</sup>

## Transcription and translation

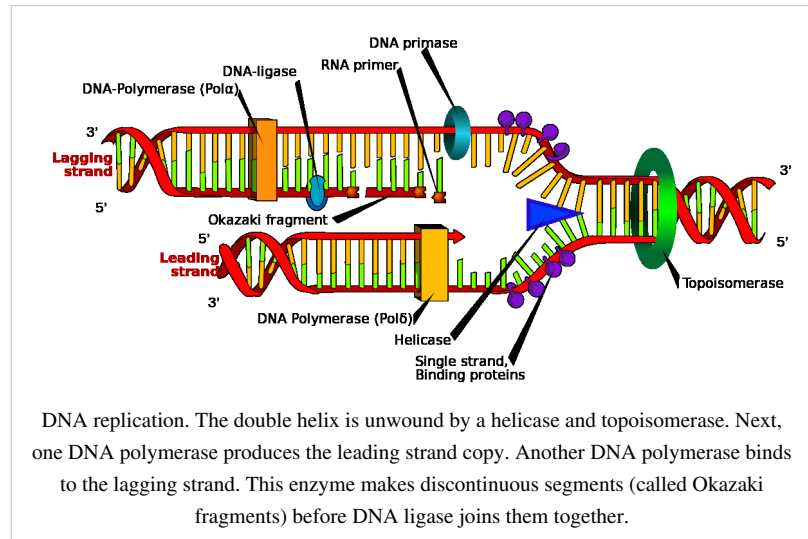
A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism. Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code. The genetic code consists of three-letter 'words' called *codons* formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries amino acids. Since there are 4 bases in 3-letter combinations, there are 64 possible codons ( $4^3$  combinations). These encode the twenty standard amino acids, giving most amino acids more than one possible codon. There are also three 'stop' or 'nonsense' codons signifying the end of the coding region; these are the TAA, TGA and TAG codons.

## Replication

Cell division is essential for an organism to grow, but, when a cell divides, it must replicate the DNA in its genome so that the two daughter cells have the same genetic information as their parent. The double-stranded structure of DNA provides a simple mechanism for DNA replication. Here, the two strands are separated and then each strand's complementary DNA sequence is recreated by an enzyme called DNA polymerase. This enzyme makes the

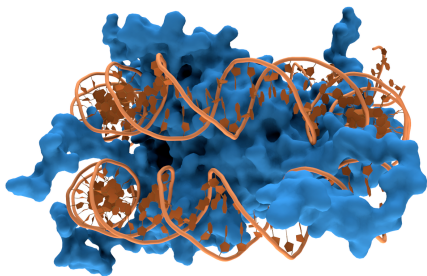
complementary strand by finding the correct base through complementary base pairing, and bonding it onto the original strand. As DNA polymerases can only extend a DNA strand in a 5' to 3' direction, different mechanisms are used to copy the antiparallel strands of the double helix.<sup>[97]</sup> In this way, the base on the old strand dictates which base appears on the new strand, and the cell ends up with a perfect copy of its DNA.



## Interactions with proteins

All the functions of DNA depend on interactions with proteins. These protein interactions can be non-specific, or the protein can bind specifically to a single DNA sequence. Enzymes can also bind to DNA and of these, the polymerases that copy the DNA base sequence in transcription and DNA replication are particularly important.

## DNA-binding proteins



Interaction of DNA (shown in orange) with histones (shown in blue). These proteins' basic amino acids bind to the acidic phosphate groups on DNA.

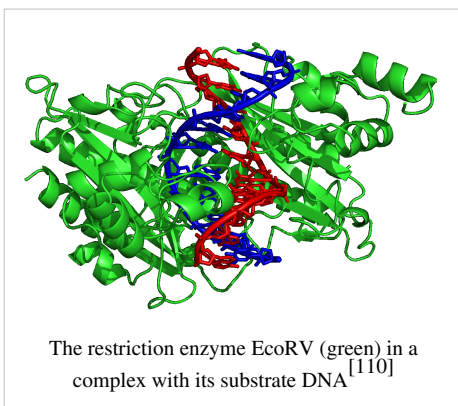
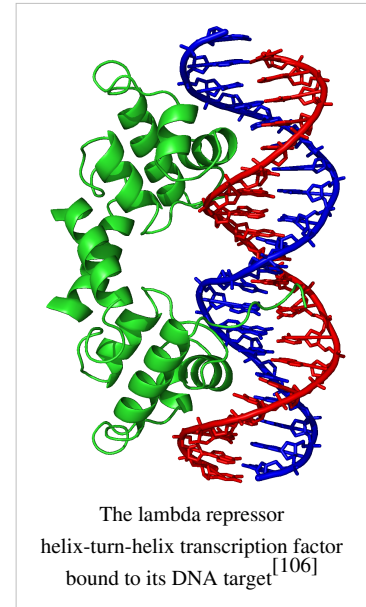
Structural proteins that bind DNA are well-understood examples of non-specific DNA-protein interactions. Within chromosomes, DNA is held in complexes with structural proteins. These proteins organize the DNA into a compact structure called chromatin. In eukaryotes this structure involves DNA binding to a complex of small basic proteins called histones, while in prokaryotes multiple types of proteins are involved.<sup>[98][99]</sup> The histones form a disk-shaped complex called a nucleosome, which contains two complete turns of double-stranded DNA wrapped around its surface. These non-specific interactions are formed through basic residues in the histones making ionic bonds to the acidic sugar-phosphate backbone of the DNA, and are therefore largely independent of the base sequence.<sup>[100]</sup> Chemical modifications of these basic amino acid residues include methylation, phosphorylation and acetylation.<sup>[101]</sup> These chemical changes alter the strength of the interaction between the DNA and the histones, making the DNA more or less accessible to transcription factors and changing the rate of transcription.<sup>[102]</sup> Other non-specific

DNA-binding proteins in chromatin include the high-mobility group proteins, which bind to bent or distorted DNA.<sup>[103]</sup> These proteins are important in bending arrays of nucleosomes and arranging them into the larger structures that make up chromosomes.<sup>[104]</sup>

A distinct group of DNA-binding proteins are the DNA-binding proteins that specifically bind single-stranded DNA. In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination and DNA repair.<sup>[105]</sup> These binding proteins seem to stabilize single-stranded DNA and protect it from forming stem-loops or being degraded by nucleases.

In contrast, other proteins have evolved to bind to particular DNA sequences. The most intensively studied of these are the various transcription factors, which are proteins that regulate transcription. Each transcription factor binds to one particular set of DNA sequences and activates or inhibits the transcription of genes that have these sequences close to their promoters. The transcription factors do this in two ways. Firstly, they can bind the RNA polymerase responsible for transcription, either directly or through other mediator proteins; this locates the polymerase at the promoter and allows it to begin transcription.<sup>[107]</sup> Alternatively, transcription factors can bind enzymes that modify the histones at the promoter; this will change the accessibility of the DNA template to the polymerase.<sup>[108]</sup>

As these DNA targets can occur throughout an organism's genome, changes in the activity of one type of transcription factor can affect thousands of genes.<sup>[109]</sup> Consequently, these proteins are often the targets of the signal transduction processes that control responses to environmental changes or cellular differentiation and development. The specificity of these transcription factors' interactions with DNA come from the proteins making multiple contacts to the edges of the DNA bases, allowing them to "read" the DNA sequence. Most of these base-interactions are made in the major groove, where the bases are most accessible.<sup>[25]</sup>



## DNA-modifying enzymes

### Nucleases and ligases

Nucleases are enzymes that cut DNA strands by catalyzing the hydrolysis of the phosphodiester bonds. Nucleases that hydrolyse nucleotides from the ends of DNA strands are called exonucleases, while endonucleases cut within strands. The most frequently used nucleases in molecular biology are the restriction endonucleases, which cut DNA at specific sequences. For instance, the EcoRV enzyme shown to the left recognizes the 6-base sequence 5'-GATATC-3' and makes a cut at the vertical line. In nature, these enzymes protect bacteria against phage infection by digesting the phage DNA when it enters the bacterial cell, acting as part of the restriction modification system.<sup>[111]</sup> In technology, these sequence-specific nucleases are used in molecular cloning and DNA fingerprinting.

Enzymes called DNA ligases can rejoin cut or broken DNA strands.<sup>[112]</sup> Ligases are particularly important in lagging strand DNA replication, as they join together the short segments of DNA produced at the replication fork into a complete copy of the DNA template. They are also used in DNA repair and genetic recombination.<sup>[112]</sup>

### Topoisomerases and helicases

Topoisomerases are enzymes with both nuclease and ligase activity. These proteins change the amount of supercoiling in DNA. Some of these enzymes work by cutting the DNA helix and allowing one section to rotate, thereby reducing its level of supercoiling; the enzyme then seals the DNA break.<sup>[37]</sup> Other types of these enzymes are capable of cutting one DNA helix and then passing a second strand of DNA through this break, before rejoining the helix.<sup>[113]</sup> Topoisomerases are required for many processes involving DNA, such as DNA replication and transcription.<sup>[38]</sup>

Helicases are proteins that are a type of molecular motor. They use the chemical energy in nucleoside triphosphates, predominantly ATP, to break hydrogen bonds between bases and unwind the DNA double helix into single strands.<sup>[114]</sup> These enzymes are essential for most processes where enzymes need to access the DNA bases.

### Polymerases

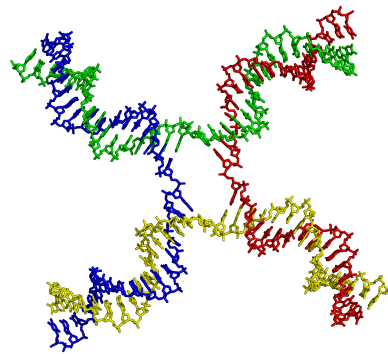
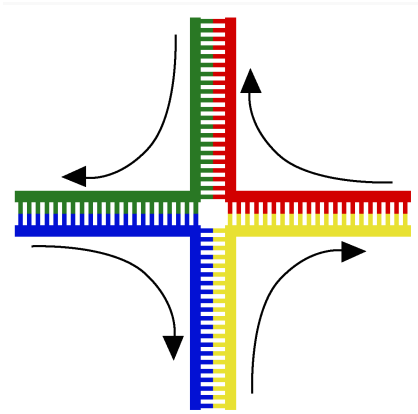
Polymerases are enzymes that synthesize polynucleotide chains from nucleoside triphosphates. The sequence of their products are copies of existing polynucleotide chains – which are called *templates*. These enzymes function by adding nucleotides onto the 3' hydroxyl group of the previous nucleotide in a DNA strand. As a consequence, all polymerases work in a 5' to 3' direction.<sup>[115]</sup> In the active site of these enzymes, the incoming nucleoside triphosphate base-pairs to the template: this allows polymerases to accurately synthesize the complementary strand of their template. Polymerases are classified according to the type of template that they use.

In DNA replication, a DNA-dependent DNA polymerase makes a copy of a DNA sequence. Accuracy is vital in this process, so many of these polymerases have a proofreading activity. Here, the polymerase recognizes the occasional mistakes in the synthesis reaction by the lack of base pairing between the mismatched nucleotides. If a mismatch is detected, a 3' to 5' exonuclease activity is activated and the incorrect base removed.<sup>[116]</sup> In most organisms, DNA polymerases function in a large complex called the replisome that contains multiple accessory subunits, such as the DNA clamp or helicases.<sup>[117]</sup>

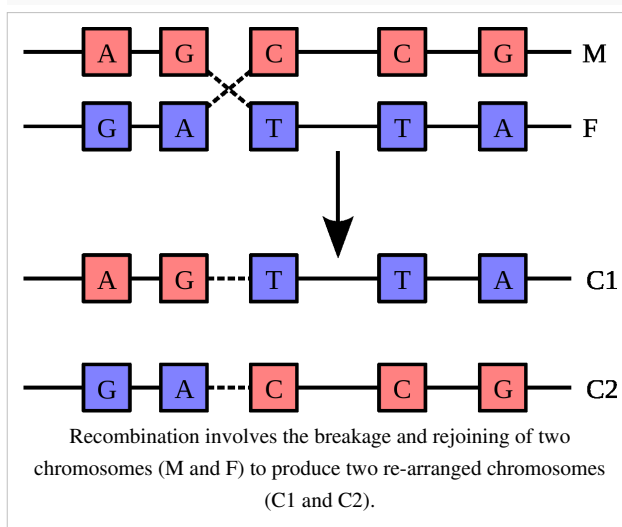
RNA-dependent DNA polymerases are a specialized class of polymerases that copy the sequence of an RNA strand into DNA. They include reverse transcriptase, which is a viral enzyme involved in the infection of cells by retroviruses, and telomerase, which is required for the replication of telomeres.<sup>[55][118]</sup> Telomerase is an unusual polymerase because it contains its own RNA template as part of its structure.<sup>[56]</sup>

Transcription is carried out by a DNA-dependent RNA polymerase that copies the sequence of a DNA strand into RNA. To begin transcribing a gene, the RNA polymerase binds to a sequence of DNA called a promoter and separates the DNA strands. It then copies the gene sequence into a messenger RNA transcript until it reaches a region of DNA called the terminator, where it halts and detaches from the DNA. As with human DNA-dependent DNA polymerases, RNA polymerase II, the enzyme that transcribes most of the genes in the human genome, operates as part of a large protein complex with multiple regulatory and accessory subunits.<sup>[119]</sup>

## Genetic recombination



Structure of the Holliday junction intermediate in genetic recombination. The four separate DNA strands are coloured red, blue, green and yellow.<sup>[120]</sup>



A DNA helix usually does not interact with other segments of DNA, and in human cells the different chromosomes even occupy separate areas in the nucleus called "chromosome territories".<sup>[121]</sup> This physical separation of different chromosomes is important for the ability of DNA to function as a stable repository for information, as one of the few times chromosomes interact is during chromosomal crossover when they recombine. Chromosomal crossover is when two DNA helices break, swap a section and then rejoin.

Recombination allows chromosomes to exchange genetic information and produces new combinations of genes, which increases the efficiency of natural selection and can be important in the rapid evolution of

new proteins.<sup>[122]</sup> Genetic recombination can also be involved in DNA repair, particularly in the cell's response to double-strand breaks.<sup>[123]</sup>

The most common form of chromosomal crossover is homologous recombination, where the two chromosomes involved share very similar sequences. Non-homologous recombination can be damaging to cells, as it can produce chromosomal translocations and genetic abnormalities. The recombination reaction is catalyzed by enzymes known as recombinases, such as RAD51.<sup>[124]</sup> The first step in recombination is a double-stranded break caused by either an endonuclease or damage to the DNA.<sup>[125]</sup> A series of steps catalyzed in part by the recombinase then leads to joining of the two helices by at least one Holliday junction, in which a segment of a single strand in each helix is annealed to the complementary strand in the other helix. The Holliday junction is a tetrahedral junction structure that can be moved along the pair of chromosomes, swapping one strand for another. The recombination reaction is then halted by cleavage of the junction and re-ligation of the released DNA.<sup>[126]</sup>

## Evolution

DNA contains the genetic information that allows all modern living things to function, grow and reproduce. However, it is unclear how long in the 4-billion-year history of life DNA has performed this function, as it has been proposed that the earliest forms of life may have used RNA as their genetic material.<sup>[127][128]</sup> RNA may have acted as the central part of early cell metabolism as it can both transmit genetic information and carry out catalysis as part of ribozymes.<sup>[129]</sup> This ancient RNA world where nucleic acid would have been used for both catalysis and genetics may have influenced the evolution of the current genetic code based on four nucleotide bases. This would occur, since the number of different bases in such an organism is a trade-off between a small number of bases increasing replication accuracy and a large number of bases increasing the catalytic efficiency of ribozymes.<sup>[130]</sup>

However, there is no direct evidence of ancient genetic systems, as recovery of DNA from most fossils is impossible. This is because DNA will survive in the environment for less than one million years and slowly degrades into short fragments in solution.<sup>[131]</sup> Claims for older DNA have been made, most notably a report of the isolation of a viable bacterium from a salt crystal 250 million years old,<sup>[132]</sup> but these claims are controversial.<sup>[133][134]</sup>

On 8 August 2011, a report, based on NASA studies with meteorites found on Earth, was published suggesting building blocks of DNA (adenine, guanine and related organic molecules) may have been formed extraterrestrially in outer space.<sup>[135][136][137]</sup>

## Uses in technology

### Genetic engineering

Methods have been developed to purify DNA from organisms, such as phenol-chloroform extraction, and to manipulate it in the laboratory, such as restriction digests and the polymerase chain reaction. Modern biology and biochemistry make intensive use of these techniques in recombinant DNA technology. Recombinant DNA is a man-made DNA sequence that has been assembled from other DNA sequences. They can be transformed into organisms in the form of plasmids or in the appropriate format, by using a viral vector.<sup>[138]</sup> The genetically modified organisms produced can be used to produce products such as recombinant proteins, used in medical research,<sup>[139]</sup> or be grown in agriculture.<sup>[140][141]</sup>

### Forensics

Forensic scientists can use DNA in blood, semen, skin, saliva or hair found at a crime scene to identify a matching DNA of an individual, such as a perpetrator. This process is formally termed DNA profiling, but may also be called "genetic fingerprinting". In DNA profiling, the lengths of variable sections of repetitive DNA, such as short tandem repeats and minisatellites, are compared between people. This method is usually an extremely reliable technique for identifying a matching DNA.<sup>[142]</sup> However, identification can be complicated if the scene is contaminated with DNA from several people.<sup>[143]</sup> DNA profiling was developed in 1984 by British geneticist Sir Alec Jeffreys,<sup>[144]</sup> and first used in forensic science to convict Colin Pitchfork in the 1988 Enderby murders case.<sup>[145]</sup>

The development of forensic science, and the ability to now obtain genetic matching on minute samples of blood, skin, saliva or hair has led to a re-examination of a number of cases. Evidence can now be uncovered that was not scientifically possible at the time of the original examination. Combined with the removal of the double jeopardy law in some places, this can allow cases to be reopened where previous trials have failed to produce sufficient evidence to convince a jury. People charged with serious crimes may be required to provide a sample of DNA for matching purposes. The most obvious defence to DNA matches obtained forensically is to claim that cross-contamination of evidence has taken place. This has resulted in meticulous strict handling procedures with new cases of serious crime. DNA profiling is also used to identify victims of mass casualty incidents.<sup>[146]</sup> As well as positively identifying bodies or body parts in serious accidents, DNA profiling is being successfully used to identify individual victims in mass war graves – matching to family members.

## Bioinformatics

Bioinformatics involves the manipulation, searching, and data mining of biological data, and this includes DNA sequence data. The development of techniques to store and search DNA sequences have led to widely applied advances in computer science, especially string searching algorithms, machine learning and database theory.<sup>[147]</sup> String searching or matching algorithms, which find an occurrence of a sequence of letters inside a larger sequence of letters, were developed to search for specific sequences of nucleotides.<sup>[148]</sup> The DNA sequence may be aligned with other DNA sequences to identify homologous sequences and locate the specific mutations that make them distinct. These techniques, especially multiple sequence alignment, are used in studying phylogenetic relationships and protein function.<sup>[149]</sup> Data sets representing entire genomes' worth of DNA sequences, such as those produced by the Human Genome Project, are difficult to use without the annotations that identify the locations of genes and regulatory elements on each chromosome. Regions of DNA sequence that have the characteristic patterns associated with protein- or RNA-coding genes can be identified by gene finding algorithms, which allow researchers to predict the presence of particular gene products and their possible functions in an organism even before they have been isolated experimentally.<sup>[150]</sup> Entire genomes may also be compared which can shed light on the evolutionary history of particular organism and permit the examination of complex evolutionary events.

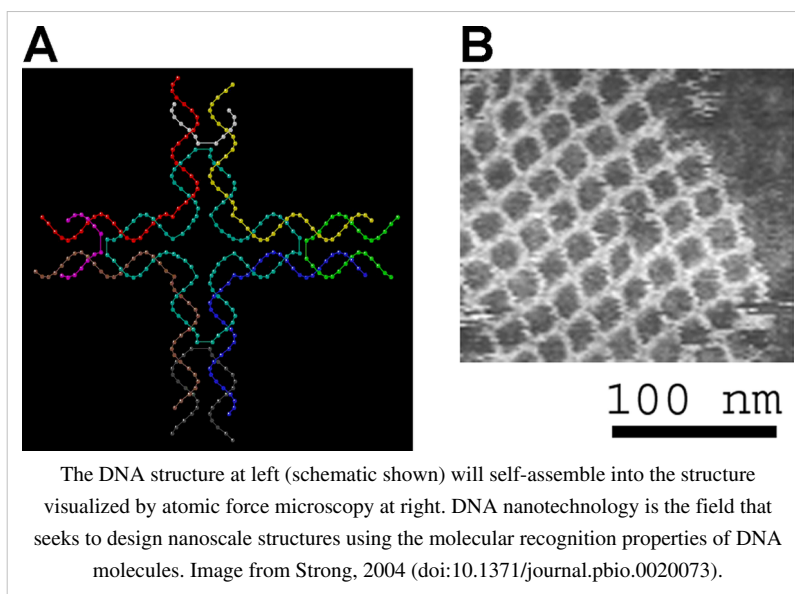
## DNA nanotechnology

DNA nanotechnology uses the unique molecular recognition properties of DNA and other nucleic acids to create self-assembling branched DNA complexes with useful properties.<sup>[151]</sup>

DNA is thus used as a structural material rather than as a carrier of biological information. This has led to the creation of two-dimensional periodic lattices (both tile-based as well as using the "DNA origami" method) as well as three-dimensional structures in the shapes of polyhedra.<sup>[152]</sup>

Nanomechanical devices and algorithmic self-assembly have also been demonstrated,<sup>[153]</sup> and

these DNA structures have been used to template the arrangement of other molecules such as gold nanoparticles and streptavidin proteins.<sup>[154]</sup>



## History and anthropology

Because DNA collects mutations over time, which are then inherited, it contains historical information, and, by comparing DNA sequences, geneticists can infer the evolutionary history of organisms, their phylogeny.<sup>[155]</sup> This field of phylogenetics is a powerful tool in evolutionary biology. If DNA sequences within a species are compared, population geneticists can learn the history of particular populations. This can be used in studies ranging from ecological genetics to anthropology; For example, DNA evidence is being used to try to identify the Ten Lost Tribes of Israel.<sup>[156][157]</sup>

DNA has also been used to look at modern family relationships, such as establishing family relationships between the descendants of Sally Hemings and Thomas Jefferson. This usage is closely related to the use of DNA in criminal



investigations detailed above. Indeed, some criminal investigations have been solved when DNA from crime scenes has matched relatives of the guilty individual.<sup>[158]</sup>

## Information storage

In a paper published in *Nature* in January, 2013, scientists from the European Bioinformatics Institute and Agilent Technologies proposed a mechanism to use DNA's ability to code information as a means of digital data storage. The group was able to encode 739 kilobytes of data into DNA code, synthesize the actual DNA, then sequence the DNA and decode the information back to its original form, with a reported 100% accuracy. The encoded information consisted of text files and audio files. A prior experiment was published in August 2012. It was conducted by researchers at Harvard University, where the text of a 54,000-word book was encoded in DNA.<sup>[159][160]</sup>

## History of DNA research

DNA was first isolated by the Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein".<sup>[161]</sup> In 1878, Albrecht Kossel isolated the non-protein component of "nuclein", nucleic acid, and later isolated its five primary nucleobases.<sup>[162]</sup> In 1919, Phoebus Levene identified the base, sugar and phosphate nucleotide unit.<sup>[163]</sup> Levene suggested that DNA consisted of a string of nucleotide units linked together through the phosphate groups. However, Levene thought the chain was short and the bases repeated in a fixed order. In 1937 William Astbury produced the first X-ray diffraction patterns that showed that DNA had a regular structure.<sup>[164]</sup>



James D. Watson and Francis Crick (right), co-originators of the double-helix model, with Maclyn McCarty (left).

In 1927 Nikolai Koltsov proposed that inherited traits would be inherited via a "giant hereditary molecule" made up of "two mirror strands that would replicate in a semi-conservative fashion using each strand as a template".<sup>[165]</sup> In 1928, Frederick Griffith discovered that traits of the "smooth" form of *Pneumococcus* could be transferred to the "rough" form of the same bacteria by mixing killed "smooth" bacteria with the live "rough" form.<sup>[166]</sup> This system provided the first clear suggestion that DNA carries genetic information—the Avery–MacLeod–McCarty experiment—when Oswald Avery, along with coworkers Colin MacLeod and Maclyn McCarty, identified DNA as the transforming principle in 1943.<sup>[167]</sup> DNA's role in heredity was confirmed in 1952, when Alfred Hershey and Martha Chase in the Hershey–Chase experiment showed that DNA is the genetic material of the T2 phage.<sup>[168]</sup>

In 1953, James D. Watson and Francis Crick suggested what is now accepted as the first correct double-helix model of DNA structure in the journal *Nature*.<sup>[5]</sup> Their double-helix, molecular model of DNA was then based on a single X-ray diffraction image (labeled as "Photo 51")<sup>[169]</sup> taken by Rosalind Franklin and Raymond Gosling in May 1952, as well as the information that the DNA bases are paired — also obtained through private communications from Erwin Chargaff in the previous years. Chargaff's rules played a very important role in establishing double-helix configurations for B-DNA as well as A-DNA.

Experimental evidence supporting the Watson and Crick model was published in a series of five articles in the same issue of *Nature*.<sup>[170]</sup> Of these, Franklin and Gosling's paper was the first publication of their own X-ray diffraction data and original analysis method that partially supported the Watson and Crick model;<sup>[41][171]</sup> this issue also contained an article on DNA structure by Maurice Wilkins and two of his colleagues, whose analysis and *in vivo* B-DNA X-ray patterns also supported the presence *in vivo* of the double-helical DNA configurations as proposed by Crick and Watson for their double-helix molecular model of DNA in the previous two pages of *Nature*.<sup>[42]</sup> In 1962, after Franklin's death, Watson, Crick, and Wilkins jointly received the Nobel Prize in Physiology or Medicine.<sup>[172]</sup>

Nobel Prizes were awarded only to living recipients at the time. A debate continues about who should receive credit for the discovery.<sup>[173]</sup>

In an influential presentation in 1957, Crick laid out the central dogma of molecular biology, which foretold the relationship between DNA, RNA, and proteins, and articulated the "adaptor hypothesis".<sup>[174]</sup> Final confirmation of the replication mechanism that was implied by the double-helical structure followed in 1958 through the Meselson–Stahl experiment.<sup>[175]</sup> Further work by Crick and coworkers showed that the genetic code was based on non-overlapping triplets of bases, called codons, allowing Har Gobind Khorana, Robert W. Holley and Marshall Warren Nirenberg to decipher the genetic code.<sup>[176]</sup> These findings represent the birth of molecular biology.

## References

- [1] Russell, Peter (2001). *iGenetics*. New York: Benjamin Cummings. ISBN 0-8053-4553-1.
- [2] Saenger, Wolfram (1984). *Principles of Nucleic Acid Structure*. New York: Springer-Verlag. ISBN 0-387-90762-9.
- [3] Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walters (2002). *Molecular Biology of the Cell; Fourth Edition* (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=mboc4.TOC&depth=2>). New York and London: Garland Science. ISBN 0-8153-3218-1. OCLC 145080076 48122761 57023651 69932405. .
- [4] Butler, John M. (2001). *Forensic DNA Typing*. Elsevier. ISBN 978-0-12-147951-0. OCLC 223032110 45406517. pp. 14–15.
- [5] Watson J.D. and Crick F.H.C. (1953). "A Structure for Deoxyribose Nucleic Acid" (<http://www.nature.com/nature/dna50/watsoncrick.pdf>) (PDF). *Nature* **171** (4356): 737–738. Bibcode 1953Natur.171..737W. doi:10.1038/171737a0. PMID 13054692. .
- [6] Mandelkern M, Elias J, Eden D, Crothers D (1981). "The dimensions of DNA in solution". *J Mol Biol* **152** (1): 153–61. doi:10.1016/0022-2836(81)90099-1. PMID 7338906.
- [7] Gregory S; Barlow, KF; McLay, KE; Kaul, R; Swarbreck, D; Dunham, A; Scott, CE; Howe, KL et al. (2006). "The DNA sequence and biological annotation of human chromosome 1". *Nature* **441** (7091): 315–21. Bibcode 2006Natur.441..315G. doi:10.1038/nature04727. PMID 16710414.
- [8] Watson J.D. and Crick F.H.C. (1953). "A Structure for Deoxyribose Nucleic Acid" (<http://www.nature.com/nature/dna50/watsoncrick.pdf>) (PDF). *Nature* **171** (4356): 737–738. Bibcode 1953Natur.171..737W. doi:10.1038/171737a0. PMID 13054692. . Retrieved 4 May 2009.
- [9] Berg J., Tymoczko J. and Stryer L. (2002) *Biochemistry*. W. H. Freeman and Company ISBN 0-7167-4955-6
- [10] Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents (<http://www.chem.qmul.ac.uk/iupac/misc/naabb.html>) IUPAC-IUB Commission on Biochemical Nomenclature (CBN). Retrieved 3 January 2006.
- [11] Ghosh A, Bansal M (2003). "A glossary of DNA structures from A to Z". *Acta Crystallogr D* **59** (4): 620–6. doi:10.1107/S0907444903003251. PMID 12657780.
- [12] Created from PDB 1D65 (<http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1D65>)
- [13] Yakovchuk P, Protozanova E, Frank-Kamenetskii MD (2006). "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix". *Nucleic Acids Res.* **34** (2): 564–74. doi:10.1093/nar/gkj454. PMC 1360284. PMID 16449200.
- [14] Verma S, Eckstein F (1998). "Modified oligonucleotides: synthesis and strategy for users". *Annu. Rev. Biochem.* **67**: 99–134. doi:10.1146/annurev.biochem.67.1.99. PMID 9759484.
- [15] Kiljunen S, Hakala K, Pinta E, Huttunen S, Pluta P, Gador A, Lönnberg H, Skurnik M (2005). "Yersiniophage phiR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine". *Microbiology* **151** (12): 4093–4102.
- [16] Simpson L (1998) A base called J. *Proc Natl Acad Sci USA* **95**(5):2037-2038
- [17] Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* **62**:235-251
- [18] Cross M, Kieft R, Sabatini R, Wilm M, de Kort M, van der Marel GA, van Leeuwen F, Borst P (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J* **18**(22):6573-6581
- [19] DiPaolo C, Kieft R, Cross M, Sabatini R (2005) Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Mol Cell* **17**(3):441-451
- [20] Vainio S, Genest PA, ter Riet B, van Luenen H, Borst P (2009) Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. *Mol Biochem Parasitol* **164**(2):157-161
- [21] Iyer LM, Tahiliani M, Rao A, Aravind L (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**(11):1698-1710
- [22] van Luenen HG, Farris C, Jan S, Genest PA, Tripathi P, Velds A, Kerkhoven RM, Nieuwland M, Haydock A, Ramasamy G, Vainio S, Heidebrecht T, Perrakis A, Pagie L, van Steensel B, Myler PJ, Borst P. Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell* **150**(5):909-921 doi: 10.1016/j.cell.2012.07.030
- [23] Hazelbaker DZ, Buratowski S (2012) Transcription: base J blocks the way. *Curr Biol* **22**(22):R960-2. doi: 10.1016/j.cub.2012.10.010
- [24] Wing R, Drew H, Takano T, Broka C, Tanaka S, Itakura K, Dickerson R (1980). "Crystal structure analysis of a complete turn of B-DNA". *Nature* **287** (5784): 755–8. Bibcode 1980Natur.287..755W. doi:10.1038/287755a0. PMID 7432492.
- [25] Pabo C, Sauer R (1984). "Protein-DNA recognition". *Annu Rev Biochem* **53**: 293–321. doi:10.1146/annurev.bi.53.070184.001453. PMID 6236744.

- [26] Clausen-Schaumann H, Rief M, Tolksdorf C, Gaub H (2000). "Mechanical stability of single DNA molecules". *Biophys J* **78** (4): 1997–2007. Bibcode 2000BpJ....78.1997C. doi:10.1016/S0006-3495(00)76747-6. PMC 1300792. PMID 10733978.
- [27] Chalikian T, Völker J, Plum G, Breslauer K (1999). "A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques". *Proc Natl Acad Sci USA* **96** (14): 7853–8. Bibcode 1999PNAS...96.7853C. doi:10.1073/pnas.96.14.7853. PMC 22151. PMID 10393911.
- [28] deHaseth P, Helmann J (1995). "Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA". *Mol Microbiol* **16** (5): 817–24. doi:10.1111/j.1365-2958.1995.tb02309.x. PMID 7476180.
- [29] Isaksson J, Acharya S, Barman J, Cheruku P, Chattopadhyaya J (2004). "Single-stranded adenine-rich DNA and RNA retain structural characteristics of their respective double-stranded conformations and show directional differences in stacking pattern". *Biochemistry* **43** (51): 15996–6010. doi:10.1021/bi048221v. PMID 15609994.
- [30] Designation of the two strands of DNA (<http://www.chem.qmul.ac.uk/iubmb/newsletter/misc/DNA.html>) JCBN/NC-IUB Newsletter 1989. Retrieved 7 May 2008
- [31] Hüttenhofer A, Schattner P, Polacek N (2005). "Non-coding RNAs: hope or hype?". *Trends Genet* **21** (5): 289–97. doi:10.1016/j.tig.2005.03.007. PMID 15851066.
- [32] Munroe S (2004). "Diversity of antisense regulation in eukaryotes: multiple mechanisms, emerging patterns". *J Cell Biochem* **93** (4): 664–71. doi:10.1002/jcb.20252. PMID 15389973.
- [33] Makalowska I, Lin C, Makalowski W (2005). "Overlapping genes in vertebrate genomes". *Comput Biol Chem* **29** (1): 1–12. doi:10.1016/j.compbiolchem.2004.12.006. PMID 15680581.
- [34] Johnson Z, Chisholm S (2004). "Properties of overlapping genes are conserved across microbial genomes". *Genome Res* **14** (11): 2268–72. doi:10.1101/gr.2433104. PMC 525685. PMID 15520290.
- [35] Lamb R, Horvath C (1991). "Diversity of coding strategies in influenza viruses". *Trends Genet* **7** (8): 261–6. doi:10.1016/0168-9525(91)90326-L. PMID 1771674.
- [36] Benham C, Mielke S (2005). "DNA mechanics". *Annu Rev Biomed Eng* **7**: 21–53. doi:10.1146/annurev.bioeng.6.062403.132016. PMID 16004565.
- [37] Champoux J (2001). "DNA topoisomerases: structure, function, and mechanism". *Annu Rev Biochem* **70**: 369–413. doi:10.1146/annurev.biochem.70.1.369. PMID 11395412.
- [38] Wang J (2002). "Cellular roles of DNA topoisomerases: a molecular perspective". *Nat Rev Mol Cell Biol* **3** (6): 430–40. doi:10.1038/nrm831. PMID 12042765.
- [39] Basu H, Feuerstein B, Zurling D, Shafer R, Marton L (1988). "Recognition of Z-RNA and Z-DNA determinants by polyamines in solution: experimental and theoretical studies". *J Biomol Struct Dyn* **6** (2): 299–309. PMID 2482766.
- [40] Franklin RE, Gosling RG (6 March 1953). "The Structure of Sodium Thymonucleate Fibres I. The Influence of Water Content" ([http://hekto.med.unc.edu:8080/CARTER/carter\\_WWW/Bioch\\_134/PDF\\_files/Franklin\\_Gosling.pdf](http://hekto.med.unc.edu:8080/CARTER/carter_WWW/Bioch_134/PDF_files/Franklin_Gosling.pdf)). *Acta Crystallogr* **6** (8–9): 673–7. doi:10.1107/S0365110X53001939. .  
Franklin RE, Gosling RG (1953). "The structure of sodium thymonucleate fibres. II. The cylindrically symmetrical Patterson function". *Acta Crystallogr* **6** (8–9): 678–85. doi:10.1107/S0365110X53001940.
- [41] Franklin, Rosalind and Gosling, Raymond (1953). "Molecular Configuration in Sodium Thymonucleate. Franklin R. and Gosling R.G" (<http://www.nature.com/nature/dna50/franklingosling.pdf>) (PDF). *Nature* **171** (4356): 740–1. Bibcode 1953Natur.171..740F. doi:10.1038/171740a0. PMID 13054694. .
- [42] Wilkins M.H.F., A.R. Stokes A.R. & Wilson, H.R. (1953). "Molecular Structure of Deoxyribose Nucleic Acids" (<http://www.nature.com/nature/dna50/wilkins.pdf>) (PDF). *Nature* **171** (4356): 738–740. Bibcode 1953Natur.171..738W. doi:10.1038/171738a0. PMID 13054693. .
- [43] Leslie AG, Arnott S, Chandrasekaran R, Ratliff RL (1980). "Polymorphism of DNA double helices". *J. Mol. Biol.* **143** (1): 49–72. doi:10.1016/0022-2836(80)90124-2. PMID 7441761.
- [44] Baianu, I.C. (1980). "Structural Order and Partial Disorder in Biological systems". *Bull. Math. Biol.* **42** (4): 137–141. <http://cogprints.org/3822/>
- [45] Hosemann R., Bagchi R.N., *Direct analysis of diffraction by matter*, North-Holland Publ., Amsterdam – New York, 1962.
- [46] Baianu, I.C. (1978). "X-ray scattering by partially disordered membrane systems". *Acta Crystallogr A* **34** (5): 751–753. Bibcode 1978AcCrA..34..751B. doi:10.1107/S0567739478001540.
- [47] Wahl M, Sundaralingam M (1997). "Crystal structures of A-DNA duplexes". *Biopolymers* **44** (1): 45–63. doi:10.1002/(SICI)1097-0282(1997)44:1<45::AID-BIP4>3.0.CO;2-#. PMID 9097733.
- [48] Lu XJ, Shakked Z, Olson WK (2000). "A-form conformational motifs in ligand-bound DNA structures". *J. Mol. Biol.* **300** (4): 819–40. doi:10.1006/jmbi.2000.3690. PMID 10891271.
- [49] Rothenburg S, Koch-Nolte F, Haag F (2001). "DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles". *Immunol Rev* **184**: 286–98. doi:10.1034/j.1600-065x.2001.1840125.x. PMID 12086319.
- [50] Oh D, Kim Y, Rich A (2002). "Z-DNA-binding proteins can act as potent effectors of gene expression in vivo". *Proc. Natl. Acad. Sci. U.S.A.* **99** (26): 16666–71. Bibcode 2002PNAS...9916666O. doi:10.1073/pnas.262672699. PMC 139201. PMID 12486233.
- [51] "Arsenic-loving bacteria may help in hunt for alien life" (<http://www.bbc.co.uk/news/science-environment-11886943>). *BBC News*. 2 December 2010. . Retrieved 2 December 2010.

- [52] Bortman, Henry (2 December 2010). "Arsenic-Eating Bacteria Opens New Possibilities for Alien Life" (<http://www.space.com/scienceastronomy/arsenic-bacteria-alien-life-101202.html>). *Space.Com web site* (<http://www.space.com/>) (Space.com). . Retrieved 2 December 2010.
- [53] Katsnelson, Alla (2 December 2010). "Arsenic-eating microbe may redefine chemistry of life" (<http://www.nature.com/news/2010/101202/full/news.2010.645.html>). *Nature News*. doi:10.1038/news.2010.645. .
- [54] "'Arsenic-life' Bacterium Prefers Phosphorus after all" (<http://www.nature.com/news/arsenic-life-bacterium-prefers-phosphorus-after-all-1.11520>). *Nature News*. 3 October 2012. .
- [55] Greider C, Blackburn E (1985). "Identification of a specific telomere terminal transferase activity in Tetrahymena extracts". *Cell* **43** (2 Pt 1): 405–13. doi:10.1016/0092-8674(85)90170-9. PMID 3907856.
- [56] Nugent C, Lundblad V (1998). "The telomerase reverse transcriptase: components and regulation". *Genes Dev* **12** (8): 1073–85. doi:10.1101/gad.12.8.1073. PMID 9553037.
- [57] Wright W, Tesmer V, Huffman K, Levene S, Shay J (1997). "Normal human chromosomes have long G-rich telomeric overhangs at one end". *Genes Dev* **11** (21): 2801–9. doi:10.1101/gad.11.21.2801. PMC 316649. PMID 9353250.
- [58] Created from NDB UD0017 (<http://ndbserver.rutgers.edu/atlas/xray/structures/U/ud0017/ud0017.html>)
- [59] Burge S, Parkinson G, Hazel P, Todd A, Neidle S (2006). "Quadruplex DNA: sequence, topology and structure". *Nucleic Acids Res* **34** (19): 5402–15. doi:10.1093/nar/gkl655. PMC 1636468. PMID 17012276.
- [60] Parkinson G, Lee M, Neidle S (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA". *Nature* **417** (6891): 876–80. doi:10.1038/nature755. PMID 12050675.
- [61] Griffith J, Comeau L, Rosenfield S, Stansel R, Bianchi A, Moss H, de Lange T (1999). "Mammalian telomeres end in a large duplex loop". *Cell* **97** (4): 503–14. doi:10.1016/S0092-8674(00)80760-6. PMID 10338214.
- [62] Seeman NC (2005). "DNA enables nanoscale control of the structure of matter". *Q. Rev. Biophys.* **38** (4): 363–71. doi:10.1017/S0033583505004087. PMID 16515737.
- [63] Painter PC, Mosher LE, Rhoads C (1982). "Low-frequency modes in the Raman spectra of proteins". *Biopolymers* **21** (7): 1469–72. doi:10.1002/bip.360210715. PMID 7115900.
- [64] Urabe H, Tominaga Y, Kubota K (1983). "Experimental evidence of collective vibrations in DNA double helix (Raman spectroscopy)". *Journal of Chemical Physics* **78** (10): 5937–5939. Bibcode 1983JChPh..78.5937U. doi:10.1063/1.444600.
- [65] Chou KC (1984). "Low-frequency vibrations of DNA molecules". *Biochem. J.* **221** (1): 27–31. PMC 1143999. PMID 6466317.
- [66] Chou KC, Maggiora GM, Mao B (1989). "Quasi-continuum models of twist-like and accordion-like low-frequency motions in DNA". *Biophys. J.* **56** (2): 295–305. Bibcode 1989BpJ....56..295C. doi:10.1016/S0006-3495(89)82676-1. PMC 1280479. PMID 2775828.
- [67] Hu Q, Rosenfeld MG. (2012) Epigenetic regulation of human embryonic stem cells. *Front Genet.* 3:238. doi: 10.3389/fgene.2012.00238. PMID 23133442
- [68] Klose R, Bird A (2006). "Genomic DNA methylation: the mark and its mediators". *Trends Biochem Sci* **31** (2): 89–97. doi:10.1016/j.tibs.2005.12.008. PMID 16403636.
- [69] Bird A (2002). "DNA methylation patterns and epigenetic memory". *Genes Dev* **16** (1): 6–21. doi:10.1101/gad.947102. PMID 11782440.
- [70] Walsh C, Xu G (2006). "Cytosine methylation and DNA repair". *Curr Top Microbiol Immunol.* Current Topics in Microbiology and Immunology **301**: 283–315. doi:10.1007/3-540-31390-7\_11. ISBN 3-540-29114-8. PMID 16570853.
- [71] Kriaucionis S, Heintz N (2009). "The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain". *Science* **324** (5929): 929–30. Bibcode 2009Sci...324..929K. doi:10.1126/science.1169786. PMC 3263819. PMID 19372393.
- [72] Ratel D, Ravanat J, Berger F, Wion D (2006). "N6-methyladenine: the other methylated base of DNA". *BioEssays* **28** (3): 309–15. doi:10.1002/bies.20342. PMC 2754416. PMID 16479578.
- [73] Gommers-Ampt J, Van Leeuwen F, de Beer A, Vliegthart J, Dizdaroglu M, Kowalak J, Crain P, Borst P (1993). "beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*". *Cell* **75** (6): 1129–36. doi:10.1016/0092-8674(93)90322-H. PMID 8261512.
- [74] Created from PDB 1JDG (<http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1JDG>)
- [75] Douki T, Reynaud-Angelin A, Cadet J, Sage E (2003). "Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation". *Biochemistry* **42** (30): 9221–6. doi:10.1021/bi034593c. PMID 12885257.
- [76] Cadet J, Delatour T, Douki T, Gasparutto D, Pouget J, Ravanat J, Sauvaigo S (1999). "Hydroxyl radicals and DNA base damage". *Mutat Res* **424** (1–2): 9–21. doi:10.1016/S0027-5107(99)00004-4. PMID 10064846.
- [77] Beckman KB, Ames BN (1997). "Oxidative decay of DNA". *J. Biol. Chem.* **272** (32): 19633–6. doi:10.1074/jbc.272.32.19633. PMID 9289489.
- [78] Valerie K, Povirk L (2003). "Regulation and mechanisms of mammalian double-strand break repair". *Oncogene* **22** (37): 5792–812. doi:10.1038/sj.onc.1206679. PMID 12947387.
- [79] Johnson, George (28 December 2010). "Unearthing Prehistoric Tumors, and Debate" (<http://www.nytimes.com/2010/12/28/health/28cancer.html>). *The New York Times*. . "If we lived long enough, sooner or later we all would get cancer."
- [80] Alberts, B, Johnson A, Lewis J, et al. (2002). "The Preventable Causes of Cancer" (<http://www.ncbi.nlm.nih.gov/books/NBK26897/>). *Molecular biology of the cell* (4th ed.). New York: Garland Science. ISBN 0-8153-4072-9. . "A certain irreducible background incidence of cancer is to be expected regardless of circumstances: mutations can never be absolutely avoided, because they are an inescapable consequence of fundamental limitations on the accuracy of DNA replication, as discussed in Chapter 5. If a human could live long enough, it is inevitable that at least one of his or her cells would eventually accumulate a set of mutations sufficient for cancer to develop."

- [81] Bernstein H, Payne CM, Bernstein C, Garewal H, Dvorak K. (2008) Cancer and aging as consequences of un-repaired DNA damage. In: *New Research on DNA Damage* (Editors: Honoka Kimura And Aoi Suzuki) Nova Science Publishers, Inc., New York, Chapter 1, pp. 1–47. ISBN 978-1-60456-581-2
- [82] Hoeijmakers JH. (2009) DNA damage, aging, and cancer. *N Engl J Med.* **361**(15):1475–1485. Review. PMID 19812404
- [83] Freitas AA, de Magalhães JP. (2011) A review and appraisal of the DNA damage theory of ageing. *Mutat Res.* **728**(1–2):12–22. Review. PMID 21600302
- [84] Ferguson L, Denny W (1991). "The genetic toxicology of acridines". *Mutat Res* **258** (2): 123–60. PMID 1881402.
- [85] Stephens T, Bunde C, Fillmore B (2000). "Mechanism of action in thalidomide teratogenesis". *Biochem Pharmacol* **59** (12): 1489–99. doi:10.1016/S0006-2952(99)00388-3. PMID 10799645.
- [86] Jeffrey A (1985). "DNA modification by chemical carcinogens". *Pharmacol Ther* **28** (2): 237–72. doi:10.1016/0163-7258(85)90013-0. PMID 3936066.
- [87] Braña M, Cacho M, Gradillas A, de Pascual-Teresa B, Ramos A (2001). "Intercalators as anticancer drugs". *Curr Pharm Des* **7** (17): 1745–80. doi:10.2174/1381612013397113. PMID 11562309.
- [88] Venter J; Adams, MD; Myers, EW; Li, PW; Mural, RJ; Sutton, GG; Smith, HO; Yandell, M et al. (2001). "The sequence of the human genome". *Science* **291** (5507): 1304–51. Bibcode 2001Sci...291.1304V. doi:10.1126/science.1058040. PMID 11181995.
- [89] Thanbichler M, Wang S, Shapiro L (2005). "The bacterial nucleoid: a highly organized and dynamic structure". *J Cell Biochem* **96** (3): 506–21. doi:10.1002/jcb.20519. PMID 15988757.
- [90] Wolfsberg T, McEntyre J, Schuler G (2001). "Guide to the draft human genome". *Nature* **409** (6822): 824–6. doi:10.1038/35057000. PMID 11236998.
- [91] Gregory T (2005). "The C-value enigma in plants and animals: a review of parallels and an appeal for partnership". *Ann Bot (Lond)* **95** (1): 133–46. doi:10.1093/aob/mci009. PMID 15596463.
- [92] The ENCODE Project Consortium (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project". *Nature* **447** (7146): 799–816. Bibcode 2007Natur.447..799B. doi:10.1038/nature05874. PMC 2212820. PMID 17571346.
- [93] Created from PDB 1MSW (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1MSW>)
- [94] Pidoux A, Allshire R (2005). "The role of heterochromatin in centromere function". *Philos Trans R Soc Lond B Biol Sci* **360** (1455): 569–79. doi:10.1098/rstb.2004.1611. PMC 1569473. PMID 15905142.
- [95] Harrison P, Hegyi H, Balasubramanian S, Luscombe N, Bertone P, Echols N, Johnson T, Gerstein M (2002). "Molecular Fossils in the Human Genome: Identification and Analysis of the Pseudogenes in Chromosomes 21 and 22". *Genome Res* **12** (2): 272–80. doi:10.1101/gr.207102. PMC 155275. PMID 11827946.
- [96] Harrison P, Gerstein M (2002). "Studying genomes through the aeons: protein families, pseudogenes and proteome evolution". *J Mol Biol* **318** (5): 1155–74. doi:10.1016/S0022-2836(02)00109-2. PMID 12083509.
- [97] Albà M (2001). "Replicative DNA polymerases". *Genome Biol* **2** (1): REVIEWS3002. doi:10.1186/gb-2001-2-1-reviews3002. PMC 150442. PMID 11178285.
- [98] Sandman K, Pereira S, Reeve J (1998). "Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome". *Cell Mol Life Sci* **54** (12): 1350–64. doi:10.1007/s000180050259. PMID 9893710.
- [99] Dame RT (2005). "The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin". *Mol. Microbiol.* **56** (4): 858–70. doi:10.1111/j.1365-2958.2005.04598.x. PMID 15853876.
- [100] Luger K, Mäder A, Richmond R, Sargent D, Richmond T (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution". *Nature* **389** (6648): 251–60. Bibcode 1997Natur.389..251L. doi:10.1038/38444. PMID 9305837.
- [101] Jenuwein T, Allis C (2001). "Translating the histone code". *Science* **293** (5532): 1074–80. doi:10.1126/science.1063127. PMID 11498575.
- [102] Ito T (2003). "Nucleosome assembly and remodelling". *Curr Top Microbiol Immunol.* Current Topics in Microbiology and Immunology **274**: 1–22. doi:10.1007/978-3-642-55747-7\_1. ISBN 978-3-540-44208-0. PMID 12596902.
- [103] Thomas J (2001). "HMG1 and 2: architectural DNA-binding proteins". *Biochem Soc Trans* **29** (Pt 4): 395–401. doi:10.1042/BST0290395. PMID 11497996.
- [104] Grosschedl R, Giese K, Pagel J (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures". *Trends Genet* **10** (3): 94–100. doi:10.1016/0168-9525(94)90232-1. PMID 8178371.
- [105] Iftode C, Daniely Y, Borowiec J (1999). "Replication protein A (RPA): the eukaryotic SSB". *Crit Rev Biochem Mol Biol* **34** (3): 141–80. doi:10.1080/10409239991209255. PMID 10473346.
- [106] Created from PDB 1LMB (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1LMB>)
- [107] Myers L, Kornberg R (2000). "Mediator of transcriptional regulation". *Annu Rev Biochem* **69**: 729–49. doi:10.1146/annurev.biochem.69.1.729. PMID 10966474.
- [108] Spiegelman B, Heinrich R (2004). "Biological control through regulated transcriptional coactivators". *Cell* **119** (2): 157–67. doi:10.1016/j.cell.2004.09.037. PMID 15479634.
- [109] Li Z, Van Calcar S, Qu C, Cavenee W, Zhang M, Ren B (2003). "A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells". *Proc Natl Acad Sci USA* **100** (14): 8164–9. Bibcode 2003PNAS..100.8164L. doi:10.1073/pnas.1332764100. PMC 166200. PMID 12808131.
- [110] Created from PDB 1RVA (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1RVA>)
- [111] Bickle T, Krüger D (1993). "Biology of DNA restriction". *Microbiol Rev* **57** (2): 434–50. PMC 372918. PMID 8336674.

- [112] Doherty A, Suh S (2000). "Structural and mechanistic conservation in DNA ligases". *Nucleic Acids Res* **28** (21): 4051–8. doi:10.1093/nar/28.21.4051. PMC 113121. PMID 11058099.
- [113] Schoeffler A, Berger J (2005). "Recent advances in understanding structure-function relationships in the type II topoisomerase mechanism". *Biochem Soc Trans* **33** (Pt 6): 1465–70. doi:10.1042/BST20051465. PMID 16246147.
- [114] Tuteja N, Tuteja R (2004). "Unraveling DNA helicases. Motif, structure, mechanism and function". *Eur J Biochem* **271** (10): 1849–63. doi:10.1111/j.1432-1033.2004.04094.x. PMID 15128295.
- [115] Joyce C, Steitz T (1995). "Polymerase structures and function: variations on a theme?". *J Bacteriol* **177** (22): 6321–9. PMC 177480. PMID 7592405.
- [116] Hubscher U, Maga G, Spadari S (2002). "Eukaryotic DNA polymerases". *Annu Rev Biochem* **71**: 133–63. doi:10.1146/annurev.biochem.71.090501.150041. PMID 12045093.
- [117] Johnson A, O'Donnell M (2005). "Cellular DNA replicases: components and dynamics at the replication fork". *Annu Rev Biochem* **74**: 283–315. doi:10.1146/annurev.biochem.73.011303.073859. PMID 15952889.
- [118] Tarrago-Litvak L, Andréola M, Nevinsky G, Sarih-Cottin L, Litvak S (1 May 1994). "The reverse transcriptase of HIV-1: from enzymology to therapeutic intervention". *FASEB J* **8** (8): 497–503. PMID 7514143.
- [119] Martínez E (2002). "Multi-protein complexes in eukaryotic gene transcription". *Plant Mol Biol* **50** (6): 925–47. doi:10.1023/A:1021258713850. PMID 12516863.
- [120] Created from PDB 1M6G (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1M6G>)
- [121] Cremer T, Cremer C (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells". *Nat Rev Genet* **2** (4): 292–301. doi:10.1038/35066075. PMID 11283701.
- [122] Pál C, Papp B, Lercher M (2006). "An integrated view of protein evolution". *Nat Rev Genet* **7** (5): 337–48. doi:10.1038/nrg1838. PMID 16619049.
- [123] O'Driscoll M, Jeggo P (2006). "The role of double-strand break repair – insights from human genetics". *Nat Rev Genet* **7** (1): 45–54. doi:10.1038/nrg1746. PMID 16369571.
- [124] Vispé S, Defais M (1997). "Mammalian Rad51 protein: a RecA homologue with pleiotropic functions". *Biochimie* **79** (9–10): 587–92. doi:10.1016/S0300-9084(97)82007-X. PMID 9466696.
- [125] Neale MJ, Keeney S (2006). "Clarifying the mechanics of DNA strand exchange in meiotic recombination". *Nature* **442** (7099): 153–8. Bibcode 2006Natur.442..153N. doi:10.1038/nature04885. PMID 16838012.
- [126] Dickman M, Ingleston S, Sedelnikova S, Rafferty J, Lloyd R, Grasby J, Hornby D (2002). "The RuvABC resolvosome". *Eur J Biochem* **269** (22): 5492–501. doi:10.1046/j.1432-1033.2002.03250.x. PMID 12423347.
- [127] Joyce G (2002). "The antiquity of RNA-based evolution". *Nature* **418** (6894): 214–21. doi:10.1038/418214a. PMID 12110897.
- [128] Orgel L (2004). "Prebiotic chemistry and the origin of the RNA world". *Crit Rev Biochem Mol Biol* **39** (2): 99–123. doi:10.1080/10409230490460765. PMID 15217990.
- [129] Davenport R (2001). "Ribozymes. Making copies in the RNA world". *Science* **292** (5520): 1278. doi:10.1126/science.292.5520.1278a. PMID 11360970.
- [130] Szathmáry E (1992). "What is the optimum size for the genetic alphabet?". *Proc Natl Acad Sci USA* **89** (7): 2614–8. Bibcode 1992PNAS..89.2614S. doi:10.1073/pnas.89.7.2614. PMC 48712. PMID 1372984.
- [131] Lindahl T (1993). "Instability and decay of the primary structure of DNA". *Nature* **362** (6422): 709–15. Bibcode 1993Natur.362..709L. doi:10.1038/362709a0. PMID 8469282.
- [132] Vreeland R, Rosenzweig W, Powers D (2000). "Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal". *Nature* **407** (6806): 897–900. doi:10.1038/35038060. PMID 11057666.
- [133] Hebsgaard M, Phillips M, Willerslev E (2005). "Geologically ancient DNA: fact or artefact?". *Trends Microbiol* **13** (5): 212–20. doi:10.1016/j.tim.2005.03.010. PMID 15866038.
- [134] Nickle D, Learn G, Rain M, Mullins J, Mittler J (2002). "Curiously modern DNA for a "250 million-year-old" bacterium". *J Mol Evol* **54** (1): 134–7. doi:10.1007/s00239-001-0025-x. PMID 11734907.
- [135] Callahan, M.P.; Smith, K.E.; Cleaves, H.J.; Ruzica, J.; Stern, J.C.; Glavin, D.P.; House, C.H.; Dworkin, J.P. (11 August 2011). "Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases" (<http://www.pnas.org/content/early/2011/08/10/1106493108>). PNAS. doi:10.1073/pnas.1106493108. . Retrieved 15 August 2011.
- [136] Steigerwald, John (8 August 2011). "NASA Researchers: DNA Building Blocks Can Be Made in Space" (<http://www.nasa.gov/topics/solarsystem/features/dna-meteorites.html>). NASA. . Retrieved 10 August 2011.
- [137] ScienceDaily Staff (9 August 2011). "DNA Building Blocks Can Be Made in Space, NASA Evidence Suggests" (<http://www.sciencedaily.com/releases/2011/08/110808220659.htm>). ScienceDaily. . Retrieved 9 August 2011.
- [138] Goff SP, Berg P (1976). "Construction of hybrid viruses containing SV40 and lambda phage DNA segments and their propagation in cultured monkey cells". *Cell* **9** (4 PT 2): 695–705. doi:10.1016/0092-8674(76)90133-1. PMID 189942.
- [139] Houdebine L (2007). "Transgenic animal models in biomedical research". *Methods Mol Biol* **360**: 163–202. doi:10.1385/1-59745-165-7:163. ISBN 1-59745-165-7. PMID 17172731.
- [140] Daniell H, Dhingra A (2002). "Multigene engineering: dawn of an exciting new era in biotechnology". *Curr Opin Biotechnol* **13** (2): 136–41. doi:10.1016/S0958-1669(02)00297-5. PMID 11950565.
- [141] Job D (2002). "Plant biotechnology in agriculture". *Biochimie* **84** (11): 1105–10. doi:10.1016/S0300-9084(02)00013-5. PMID 12595138.

- [142] Collins A, Morton N (1994). "Likelihood ratios for DNA identification". *Proc Natl Acad Sci USA* **91** (13): 6007–11. Bibcode 1994PNAS...91.6007C. doi:10.1073/pnas.91.13.6007. PMC 44126. PMID 8016106.
- [143] Weir B, Triggs C, Starling L, Stowell L, Walsh K, Buckleton J (1997). "Interpreting DNA mixtures". *J Forensic Sci* **42** (2): 213–22. PMID 9068179.
- [144] Jeffreys A, Wilson V, Thein S (1985). "Individual-specific 'fingerprints' of human DNA". *Nature* **316** (6023): 76–9. Bibcode 1985Natur.316...76J. doi:10.1038/316076a0. PMID 2989708.
- [145] Colin Pitchfork — first murder conviction on DNA evidence also clears the prime suspect ([http://web.archive.org/web/20061214004903/http://www.forensic.gov.uk/forensic\\_t/inside/news/list\\_casefiles.php?case=1](http://web.archive.org/web/20061214004903/http://www.forensic.gov.uk/forensic_t/inside/news/list_casefiles.php?case=1)) Forensic Science Service Accessed 23 December 2006
- [146] "DNA Identification in Mass Fatality Incidents" (<http://massfatality.dna.gov/Introduction/>). National Institute of Justice. September 2006. .
- [147] Baldi, Pierre; Brunak, Soren (2001). *Bioinformatics: The Machine Learning Approach*. MIT Press. ISBN 978-0-262-02506-5. OCLC 45951728.
- [148] Gusfield, Dan. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 15 January 1997. ISBN 978-0-521-58519-4.
- [149] Sjölander K (2004). "Phylogenomic inference of protein molecular function: advances and challenges". *Bioinformatics* **20** (2): 170–9. doi:10.1093/bioinformatics/bth021. PMID 14734307.
- [150] Mount DM (2004). *Bioinformatics: Sequence and Genome Analysis* (2 ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. ISBN 0-87969-712-1. OCLC 55106399.
- [151] Rothmund PW (2006). "Folding DNA to create nanoscale shapes and patterns". *Nature* **440** (7082): 297–302. Bibcode 2006Natur.440..297R. doi:10.1038/nature04586. PMID 16541064.
- [152] Andersen ES, Dong M, Nielsen MM (2009). "Self-assembly of a nanoscale DNA box with a controllable lid". *Nature* **459** (7243): 73–6. Bibcode 2009Natur.459...73A. doi:10.1038/nature07971. PMID 19424153.
- [153] Ishitsuka Y, Ha T (2009). "DNA nanotechnology: a nanomachine goes live". *Nat Nanotechnol* **4** (5): 281–2. Bibcode 2009NatNa...4..281I. doi:10.1038/nnano.2009.101. PMID 19421208.
- [154] Aldaye FA, Palmer AL, Sleiman HF (2008). "Assembling materials with DNA as the guide". *Science* **321** (5897): 1795–9. Bibcode 2008Sci...321.1795A. doi:10.1126/science.1154533. PMID 18818351.
- [155] Wray G; Martindale, Mark Q. (2002). "Dating branches on the Tree of Life using DNA". *Genome Biol* **3** (1): REVIEWS0001. doi:10.1046/j.1525-142X.1999.99010.x. PMC 150454. PMID 11806830.
- [156] *Lost Tribes of Israel*, NOVA, PBS airdate: 22 February 2000. Transcript available from PBS.org (<http://www.pbs.org/wgbh/nowa/transcripts/2706israel.html>). Retrieved 4 March 2006.
- [157] Kleiman, Yaakov. "The Cohanim/DNA Connection: The fascinating story of how DNA studies confirm an ancient biblical tradition". ([http://www.aish.com/societywork/sciencenature/the\\_cohanim\\_-\\_dna\\_connection.asp](http://www.aish.com/societywork/sciencenature/the_cohanim_-_dna_connection.asp)) *aish.com* (13 January 2000). Retrieved 4 March 2006.
- [158] Bhattacharya, Shaoni. "Killer convicted thanks to relative's DNA". (<http://www.newscientist.com/article.ns?id=dn4908>) *newscientist.com* (20 April 2004). Retrieved 22 December 06.
- [159] Goldman, Nick; Bertone, Paul; Chen, Siyuan; Dessimoz, Christophe; LeProust, Emily M.; Sipos, Botond; Birney, Ewan (23 January 2013). "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA" (<http://www.nature.com/nature/journal/vaop/ncurrent/full/nature11875.html>). *Nature*. doi:10.1038/nature11875. . Retrieved 24 January 2013.
- [160] Naik, Gautam (24 January 2013). "Storing Digital Data in DNA" (<http://online.wsj.com/article/SB10001424127887324539304578259883507543150.html>). *Wall Street Journal*. . Retrieved 24 January 2013.
- [161] Dahm R (2008). "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research". *Hum. Genet.* **122** (6): 565–81. doi:10.1007/s00439-007-0433-0. PMID 17901982.
- [162] Jones, Mary Ellen (September 1953). "Albrecht Kossel, A Biographical Sketch". *Yale Journal of Biology and Medicine* (National Center for Biotechnology Information) **26** (1): 80–97. PMC 2599350. PMID 13103145.
- [163] Levene P, (1 December 1919). "The structure of yeast nucleic acid". *J Biol Chem* **40** (2): 415–24.
- [164] Astbury W, (1947). "Nucleic acid". *Symp. SOC. Exp. Biol.* **1** (66).
- [165] Valery N. Soyfer (2001). "The consequences of political dictatorship for Russian science". *Nature Reviews Genetics* **2** (9): 723–729. doi:10.1038/35088598. PMID 11533721.
- [166] Lorenz MG, Wackernagel W (1994). "Bacterial gene transfer by natural genetic transformation in the environment". *Microbiol. Rev.* **58** (3): 563–602. PMC 372978. PMID 7968924.
- [167] Avery O, MacLeod C, McCarty M (1944). "STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III". *J Exp Med* **79** (2): 137–158. doi:10.1084/jem.79.2.137. PMC 2135445. PMID 19871359.
- [168] Hershey A, Chase M (1952). "INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE". *J Gen Physiol* **36** (1): 39–56. doi:10.1085/jgp.36.1.39. PMC 2147348. PMID 12981234.
- [169] The B-DNA X-ray pattern on the right of this linked image (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/pictures/sci9.001.5.html>) was obtained by Rosalind Franklin and Raymond Gosling in May 1952 at high hydration levels of DNA and it has



- been labeled as "Photo 51"
- [170] Nature Archives Double Helix of DNA: 50 Years (<http://www.nature.com/nature/dna50/archive.html>)
- [171] "Original X-ray diffraction image" (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/pictures/franklin-typeBphoto.html>). Osulibrary.oregonstate.edu. . Retrieved 6 February 2011.
- [172] The Nobel Prize in Physiology or Medicine 1962 ([http://nobelprize.org/nobel\\_prizes/medicine/laureates/1962/](http://nobelprize.org/nobel_prizes/medicine/laureates/1962/)) Nobelprize .org Accessed 22 December 06
- [173] Brenda Maddox (23 January 2003). "The double helix and the 'wronged heroine'" ([http://www.biomath.nyu.edu/index/course/hw\\_articles/nature4.pdf](http://www.biomath.nyu.edu/index/course/hw_articles/nature4.pdf)) (PDF). *Nature* **421** (6921): 407–408. doi:10.1038/nature01399. PMID 12540909. .
- [174] Crick, F.H.C. On degenerate templates and the adaptor hypothesis (PDF). (<http://genome.wellcome.ac.uk/assets/wtx030893.pdf>) genome.wellcome.ac.uk (Lecture, 1955). Retrieved 22 December 2006.
- [175] Meselson M, Stahl F (1958). "THE REPLICATION OF DNA IN ESCHERICHIA COLI". *Proc Natl Acad Sci USA* **44** (7): 671–82. Bibcode 1958PNAS..44..671M. doi:10.1073/pnas.44.7.671. PMC 528642. PMID 16590258.
- [176] The Nobel Prize in Physiology or Medicine 1968 ([http://nobelprize.org/nobel\\_prizes/medicine/laureates/1968/](http://nobelprize.org/nobel_prizes/medicine/laureates/1968/)) Nobelprize.org Accessed 22 December 06

## Further reading

- Berry, Andrew; Watson, James D. (2003). *DNA: the secret of life*. New York: Alfred A. Knopf. ISBN 0-375-41546-7.
- Calladine, Chris R.; Drew, Horace R.; Luisi, Ben F. and Travers, Andrew A. (2003). *Understanding DNA: the molecule & how it works*. Amsterdam: Elsevier Academic Press. ISBN 0-12-155089-3.
- Dennis, Carina; Julie Clayton (2003). *50 years of DNA*. Basingstoke: Palgrave Macmillan. ISBN 1-4039-1479-6.
- Judson, Horace F. 1979. *The Eighth Day of Creation: Makers of the Revolution in Biology*. Touchstone Books, ISBN 0-671-22540-5. 2nd edition: Cold Spring Harbor Laboratory Press, 1996 paperback: ISBN 0-87969-478-5.
- Olby, Robert C. (1994). *The path to the double helix: the discovery of DNA*. New York: Dover Publications. ISBN 0-486-68117-3., first published in October 1974 by MacMillan, with foreword by Francis Crick;the definitive DNA textbook, revised in 1994 with a 9 page postscript
- Micklas, David. 2003. *DNA Science: A First Course*. Cold Spring Harbor Press: ISBN 978-0-87969-636-8
- Ridley, Matt (2006). *Francis Crick: discoverer of the genetic code*. Ashland, OH: Eminent Lives, Atlas Books. ISBN 0-06-082333-X.
- Olby, Robert C. (2009). *Francis Crick: A Biography*. Plainview, N.Y: Cold Spring Harbor Laboratory Press. ISBN 0-87969-798-9.
- Rosenfeld, Israel. 2010. *DNA: A Graphic Guide to the Molecule that Shook the World*. Columbia University Press: ISBN 978-0-231-14271-7
- Schultz, Mark and Zander Cannon. 2009. *The Stuff of Life: A Graphic Guide to Genetics and DNA*. Hill and Wang: ISBN 0-8090-8947-5
- Stent, Gunther Siegmund; Watson, James D. (1980). *The double helix: a personal account of the discovery of the structure of DNA*. New York: Norton. ISBN 0-393-95075-1.
- Watson, James D. 2004. *DNA: The Secret of Life*. Random House: ISBN 978-0-09-945184-6
- Wilkins, Maurice (2003). *The third man of the double helix the autobiography of Maurice Wilkins*. Cambridge, Eng: University Press. ISBN 0-19-860665-6.

## External links

- DNA ([http://www.dmoz.org/Science/Biology/Biochemistry\\_and\\_Molecular\\_Biology/Biomolecules/Nucleic\\_Acids/DNA/](http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Biomolecules/Nucleic_Acids/DNA/)) at the Open Directory Project
- DNA binding site prediction on protein (<http://pipe.scs.fsu.edu/displar.html>)
- DNA the Double Helix Game ([http://nobelprize.org/educational\\_games/medicine/dna\\_double\\_helix/](http://nobelprize.org/educational_games/medicine/dna_double_helix/)) From the official Nobel Prize web site
- DNA under electron microscope ([http://www.fidelitysystems.com/Unlinked\\_DNA.html](http://www.fidelitysystems.com/Unlinked_DNA.html))
- Dolan DNA Learning Center (<http://www.dnalc.org/>)
- Double Helix: 50 years of DNA (<http://www.nature.com/nature/dna50/archive.html>), *Nature*
- *Proteopedia DNA* (<http://www.proteopedia.org/wiki/index.php/DNA>)
- Double Helix 1953–2003 (<http://www.ncbe.reading.ac.uk/DNA50/>) National Centre for Biotechnology Education
- Francis Crick and James Watson talking on the BBC in 1962, 1972, and 1974 (<http://www.bbc.co.uk/bbcfour/audiointerviews/profilepages/crickwatson1.shtml>)
- Genetic Education Modules for Teachers (<http://www.genome.gov/10506718>)—*DNA from the Beginning Study Guide*
- Olby R (2003). "Quiet debut for the double helix". *Nature* **421** (6921): 402–5. doi:10.1038/nature01397. PMID 12540907.
- DNA from the Beginning (<http://www.dnafb.org/dnafb/>) Another DNA Learning Center site on DNA, genes, and heredity from Mendel to the human genome project.
- PDB Molecule of the Month *pdb23\_1* ([http://www.rcsb.org/pdb/static.do?p=education\\_discussion/molecule\\_of\\_the\\_month/pdb23\\_1.html](http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb23_1.html))
- Rosalind Franklin's contributions to the study of DNA (<http://mason.gmu.edu/~emoody/rfranklin.html>)
- The Register of Francis Crick Personal Papers 1938 – 2007 (<http://orpheus.ucsd.edu/speccoll/testing/html/mss0660a.html#abstract>) at Mandeville Special Collections Library, University of California, San Diego
- U.S. National DNA Day (<http://www.genome.gov/10506367>)—watch videos and participate in real-time chat with top scientists
- Clue to chemistry of heredity found (<http://www.nytimes.com/packages/pdf/science/dna-article.pdf>) The New York Times June 1953. First American newspaper coverage of the discovery of the DNA structure

# Computer

## Computer



A **computer** is a general purpose device that can be programmed to carry out a finite set of arithmetic or logical operations. Since a sequence of operations can be readily changed, the computer can solve more than one kind of problem.

Conventionally, a computer consists of at least one processing element, typically a central processing unit (CPU) and some form of memory. The processing element carries out arithmetic and logic operations, and a sequencing and control unit that can change the order of operations based on stored information. Peripheral devices allow information to be retrieved from an external source, and the result of operations saved and retrieved.

The first electronic digital computers were developed between 1940 and 1945 in the United Kingdom and United States. Originally they were the size of a large room, consuming as much power as several hundred modern personal computers (PCs).<sup>[1]</sup> In this era mechanical analog computers were used for military applications.

Modern computers based on integrated circuits are millions to billions of times more capable than the early machines, and occupy a fraction of the space.<sup>[2]</sup> Simple computers are small enough to fit into mobile devices, and mobile computers can be powered by small batteries. Personal computers in their various forms are icons of the Information Age and are what most people think of as "computers". However, the embedded computers found in many devices from MP3 players to fighter aircraft and from toys to industrial robots are the most numerous.

## History of computing

The first use of the word "computer" was recorded in 1613 in a book called "The yong mans gleanings" by English writer Richard Braithwait *I haue read the truest computer of Times, and the best Arithmetician that euer breathed, and he reduceth thy dayes into a short number.* It referred to a person who carried out calculations, or computations, and the word continued with the same meaning until the middle of the 20th century. From the end of the 19th century the word began to take on its more familiar meaning, a machine that carries out computations.<sup>[3]</sup>

### Limited-function early computers

The history of the modern computer begins with two separate technologies, automated calculation and programmability. However no single device can be identified as the earliest computer, partly because of the inconsistent application of that term. A few devices are worth mentioning though, like some mechanical



The Jacquard loom, on display at the Museum of Science and Industry in Manchester, England, was one of the first programmable devices.

aids to computing, which were very successful and survived for centuries until the advent of the electronic calculator, like the Sumerian abacus, designed around 2500 BC<sup>[4]</sup> of which a descendant won a speed competition against a modern desk calculating machine in Japan in 1946,<sup>[5]</sup> the slide rules, invented in the 1620s, which were carried on five Apollo space missions, including to the moon<sup>[6]</sup> and arguably the astrolabe and the Antikythera mechanism, an ancient astronomical computer built by the Greeks around 80 BC.<sup>[7]</sup> The Greek mathematician Hero of Alexandria (c. 10–70 AD) built a mechanical theater which performed a play lasting 10 minutes and was operated by a complex system of ropes and drums that might be considered to be a means of deciding which parts of the mechanism performed which actions and when.<sup>[8]</sup> This is the essence of programmability.

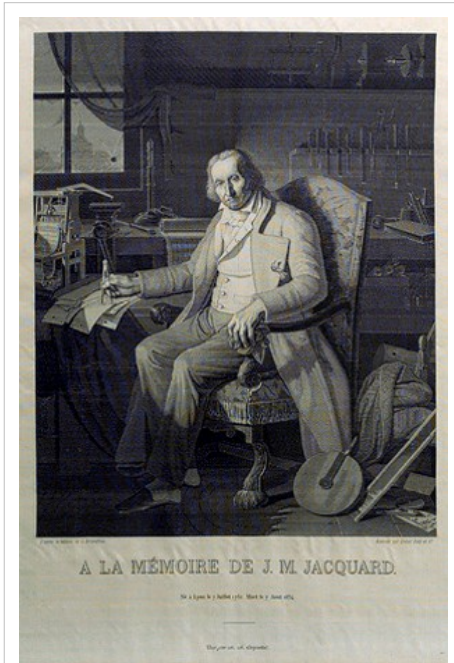
Around the end of the 10th century, the French monk Gerbert d'Aurillac brought back from Spain the drawings of a machine invented by the Moors that answered either Yes or No to the questions it was asked.<sup>[9]</sup> Again in the 13th century, the monks Albertus Magnus and Roger Bacon built talking androids without any further development (Albertus Magnus complained that he had wasted forty years of his life when Thomas Aquinas, terrified by his machine, destroyed it).<sup>[10]</sup>

In 1642, the Renaissance saw the invention of the mechanical calculator,<sup>[9]</sup> a device that could perform all four arithmetic operations without relying on human intelligence.<sup>[11]</sup> The mechanical calculator was at the root of the development of computers in two separate ways. Initially, it was in trying to develop more powerful and more flexible calculators<sup>[12]</sup> that the computer was first theorized by Charles Babbage<sup>[13][14]</sup> and then developed.<sup>[15]</sup> Secondly, development of a low-cost electronic calculator, successor to the mechanical calculator, resulted in the development by Intel<sup>[16]</sup> of the first commercially available microprocessor integrated circuit.

---

## First general-purpose computers

In 1801, Joseph Marie Jacquard made an improvement to the textile loom by introducing a series of punched paper cards as a template which allowed his loom to weave intricate patterns automatically. The resulting Jacquard loom was an important step in the development of computers because the use of punched cards to define woven patterns can be viewed as an early, albeit limited, form of programmability.



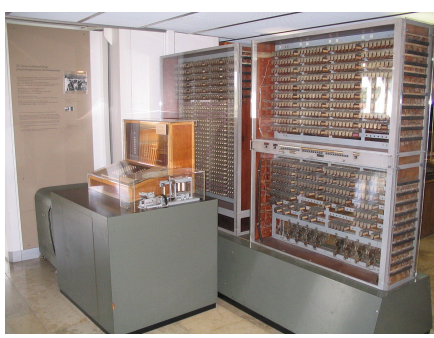
The Most Famous Image in the Early History of Computing From cave paintings to the internet HistoryofScience.com This portrait of Jacquard was woven in silk on a Jacquard loom and required 24,000 punched cards to create (1839). It was only produced to order. Charles Babbage owned one of these portraits ; it inspired him in using perforated cards in his analytical engine See: Anthony Hyman, ed., *Science and Reform: Selected Works of Charles Babbage* (Cambridge, England: Cambridge University Press, 1989), page 298. It is in the collection of the Science Museum in London, England. (Delve (2007), page 99.)

It was the fusion of automatic calculation with programmability that produced the first recognizable computers. In 1837, Charles Babbage was the first to conceptualize and design a fully programmable mechanical computer, his analytical engine.<sup>[19]</sup> Limited finances and Babbage's inability to resist tinkering with the design meant that the device was never completed—nevertheless his son, Henry Babbage, completed a simplified version of the analytical engine's computing unit (the *mill*) in 1888. He gave a successful demonstration of its use in computing tables in 1906. This machine was given to the Science museum in South Kensington in 1910.

In the late 1880s, Herman Hollerith invented the recording of data on a machine-readable medium. Earlier uses of machine-readable media had been for control, not data. "After some initial trials with paper tape, he settled on punched cards ..."<sup>[20]</sup> To process these punched cards he invented the tabulator, and the keypunch machines. These three inventions were the foundation of the modern information processing industry. Large-scale automated data processing of punched cards was performed for the 1890 United States Census by Hollerith's company, which later became the core of IBM. By the end of the 19th century a number of ideas and technologies, that would later prove useful in the realization of practical computers, had begun to appear: Boolean algebra, the vacuum tube (thermionic valve), punched cards and tape, and the teleprinter.

During the first half of the 20th century, many scientific computing needs were met by increasingly sophisticated analog computers, which used a direct mechanical or electrical model of the problem as a basis for computation. However, these were not programmable and generally lacked the versatility and accuracy of modern digital computers.

Alan Turing is widely regarded as the father of modern computer science. In 1936 Turing provided an influential formalisation of the concept of the algorithm and computation with the Turing machine, providing a blueprint for the electronic digital computer.<sup>[21]</sup> Of his role in the creation of the modern computer, *Time* magazine in naming Turing one of the 100 most influential people of the 20th century, states: "The fact remains that everyone who taps at a keyboard, opening a spreadsheet or a word-processing program, is working on an incarnation of a Turing machine".<sup>[21]</sup>



The Zuse Z3, 1941, considered the world's first working programmable, fully automatic computing machine.

The Atanasoff–Berry Computer (ABC) was the world's first electronic digital computer, albeit not programmable.<sup>[22]</sup> Atanasoff is considered to be one of the fathers of the computer.<sup>[23]</sup> Conceived in 1937 by Iowa State College physics professor John Atanasoff, and built with the assistance of graduate student Clifford Berry,<sup>[24]</sup> the machine was not programmable, being designed only to solve systems of linear equations. The computer did employ parallel computation. A 1973 court ruling in a patent dispute found that the patent for the 1946 ENIAC computer derived from the Atanasoff–Berry Computer.

The first program-controlled computer was invented by Konrad Zuse, who built the Z3, an electromechanical computing machine, in 1941.<sup>[25]</sup> The first programmable electronic computer was the Colossus, built in 1943 by Tommy Flowers.

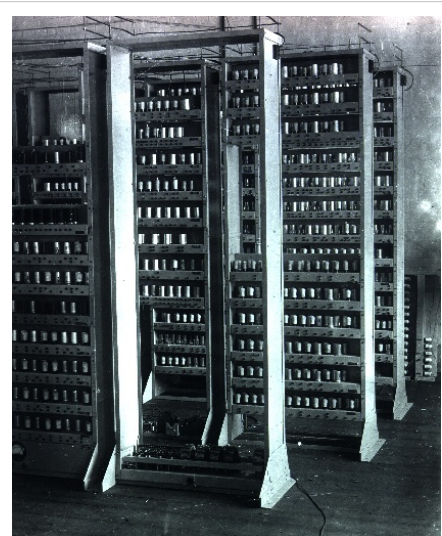
George Stibitz is internationally recognized as a father of the modern digital computer. While working at Bell Labs in November 1937, Stibitz invented and built a relay-based calculator he dubbed the "Model K" (for "kitchen table", on which he had assembled it), which was the first to use binary circuits to perform an arithmetic operation. Later models added greater sophistication including complex arithmetic and programmability.<sup>[26]</sup>

A succession of steadily more powerful and flexible computing devices were constructed in the 1930s and 1940s, gradually adding the key features that are seen in modern computers. The use of digital electronics (largely invented by Claude Shannon in 1937) and more flexible programmability were vitally important steps, but defining one point along this road as "the first digital electronic computer" is difficult.<sup>Shannon 1940</sup> Notable achievements include:

- Konrad Zuse's electromechanical "Z machines". The Z3 (1941) was the first working machine featuring binary arithmetic, including floating point arithmetic and a measure of programmability. In 1998 the Z3 was proved to be Turing complete, therefore being the world's first operational computer.<sup>[27]</sup>
- The non-programmable Atanasoff–Berry Computer (commenced in 1937, completed in 1941) which used vacuum tube based computation, binary numbers, and regenerative capacitor memory. The use of regenerative



The ENIAC, which became operational in 1946, is considered to be the first general-purpose electronic computer.



EDSAC was one of the first computers to implement the stored-program (von Neumann) architecture.



memory allowed it to be much more compact than its peers (being approximately the size of a large desk or workbench), since intermediate results could be stored and then fed back into the same set of computation elements.

- The secret British Colossus computers (1943),<sup>[28]</sup> which had limited programmability but demonstrated that a device using thousands of tubes could be reasonably reliable and electronically reprogrammable. It was used for breaking German wartime codes.
- The Harvard Mark I (1944), a large-scale electromechanical computer with limited programmability.<sup>[29]</sup>
- The U.S. Army's Ballistic Research Laboratory ENIAC (1946), which used decimal arithmetic and is sometimes called the first general purpose electronic computer (since Konrad Zuse's Z3 of 1941 used electromagnets instead of electronics). Initially, however, ENIAC had an architecture which required rewiring a plugboard to change its programming.

### Stored-program architecture

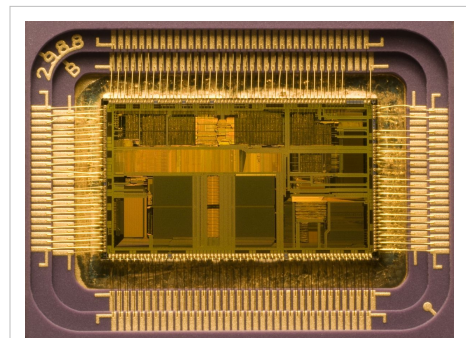
Several developers of ENIAC, recognizing its flaws, came up with a far more flexible and elegant design, which came to be known as the "stored-program architecture" or von Neumann architecture. This design was first formally described by John von Neumann in the paper *First Draft of a Report on the EDVAC*, distributed in 1945. A number of projects to develop computers based on the stored-program architecture commenced around this time, the first of which was completed in 1948 at the University of Manchester in England, the Manchester Small-Scale Experimental Machine (SSEM or "Baby"). The Electronic Delay Storage Automatic Calculator (EDSAC), completed a year after the SSEM at Cambridge University, was the first practical, non-experimental implementation of the stored-program design and was put to use immediately for research work at the university. Shortly thereafter, the machine originally described by von Neumann's paper—EDVAC—was completed but did not see full-time use for an additional two years.

Nearly all modern computers implement some form of the stored-program architecture, making it the single trait by which the word "computer" is now defined. While the technologies used in computers have changed dramatically since the first electronic, general-purpose computers of the 1940s, most still use the von Neumann architecture.

Beginning in the 1950s, Soviet scientists Sergei Sobolev and Nikolay Brusentsov conducted research on ternary computers, devices that operated on a base three numbering system of  $-1$ ,  $0$ , and  $1$  rather than the conventional binary numbering system upon which most computers are based. They designed the Setun, a functional ternary computer, at Moscow State University. The device was put into limited production in the Soviet Union, but supplanted by the more common binary architecture.

### Semiconductors and microprocessors

Computers using vacuum tubes as their electronic elements were in use throughout the 1950s, but by the 1960s they had been largely replaced by transistor-based machines, which were smaller, faster, cheaper to produce, required less power, and were more reliable. The first transistorised computer was demonstrated at the University of Manchester in 1953.<sup>[30]</sup> In the 1970s, integrated circuit technology and the subsequent creation of microprocessors, such as the Intel 4004, further decreased size and cost and further increased speed and reliability of computers. By the late 1970s, many products such as video recorders contained dedicated computers called microcontrollers, and they started to appear as a replacement to mechanical controls in domestic appliances such as washing machines. The 1980s witnessed home computers and the now ubiquitous personal computer. With the evolution of the Internet, personal computers are becoming as common as the television and the telephone in the household.



Die of an Intel 80486DX2 microprocessor (actual size: 12×6.75 mm) in its packaging



Modern smartphones are fully programmable computers in their own right, and as of 2009 may well be the most common form of such computers in existence.

## Programs

The defining feature of modern computers which distinguishes them from all other machines is that they can be programmed. That is to say that some type of instructions (the program) can be given to the computer, and it will process them. Modern computers based on the von Neumann architecture often have machine code in the form of an imperative programming language.

In practical terms, a computer program may be just a few instructions or extend to many millions of instructions, as do the programs for word processors and web browsers for example. A typical modern computer can execute billions of instructions per second (gigaflops) and rarely makes a mistake over many years of operation. Large computer programs consisting of several million instructions may take teams of programmers years to write, and due to the complexity of the task almost certainly contain errors.

## Stored program architecture

This section applies to most common RAM machine-based computers.

In most cases, computer instructions are simple: add one number to another, move some data from one location to another, send a message to some external device, etc. These instructions are read from the computer's memory and are generally carried out (executed) in the order they were given. However, there are usually specialized instructions to tell the computer to jump ahead or backwards to some other place in the program and to carry on executing from there. These are called "jump" instructions (or branches). Furthermore, jump instructions may be made to happen conditionally so that different sequences of instructions may be used depending on the result of some previous calculation or some external event. Many computers directly support subroutines by providing a type of jump that "remembers" the location it jumped from and another instruction to return to the instruction following that jump instruction.

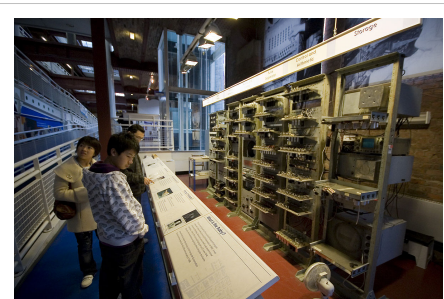
Program execution might be likened to reading a book. While a person will normally read each word and line in sequence, they may at times jump back to an earlier place in the text or skip sections that are not of interest. Similarly, a computer may sometimes go back and repeat the instructions in some section of the program over and over again until some internal condition is met. This is called the flow of control within the program and it is what allows the computer to perform tasks repeatedly without human intervention.

Comparatively, a person using a pocket calculator can perform a basic arithmetic operation such as adding two numbers with just a few button presses. But to add together all of the numbers from 1 to 1,000 would take thousands of button presses and a lot of time, with a near certainty of making a mistake. On the other hand, a computer may be programmed to do this with just a few simple instructions. For example:

```

mov No. 0, sum      ; set sum to 0
mov No. 1, num      ; set num to 1
loop: add num, sum  ; add num to sum
      add No. 1, num ; add 1 to num
      cmp num, #1000 ; compare num to 1000
      ble loop      ; if num <= 1000, go back to 'loop'
      halt         ; end of program. stop running

```



Replica of the Small-Scale Experimental Machine (SSEM), the world's first stored-program computer, at the Museum of Science and Industry in Manchester, England

Once told to run this program, the computer will perform the repetitive addition task without further human intervention. It will almost never make a mistake and a modern PC can complete the task in about a millionth of a second.<sup>[31]</sup>

## Bugs

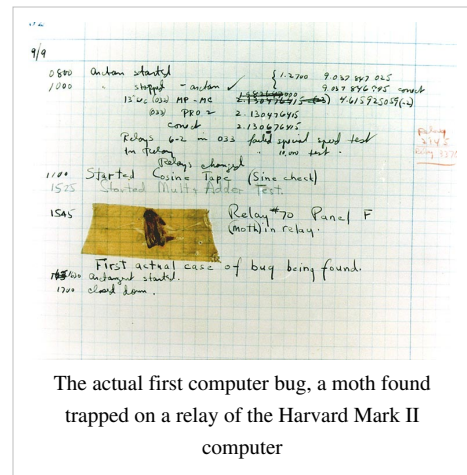
Errors in computer programs are called "bugs". They may be benign and not affect the usefulness of the program, or have only subtle effects. But in some cases they may cause the program or the entire system to "hang" – become unresponsive to input such as mouse clicks or keystrokes – to completely fail, or to crash. Otherwise benign bugs may sometimes be harnessed for malicious intent by an unscrupulous user writing an exploit, code designed to take advantage of a bug and disrupt a computer's proper execution. Bugs are usually not the fault of the computer. Since computers merely execute the instructions they are given, bugs are nearly always the result of programmer error or an oversight made in the program's design.<sup>[32]</sup>

Grace Hopper is credited for having first used the term "bugs" in computing after a dead moth was found shorting a relay in the Harvard Mark II computer in September 1947.<sup>[33]</sup>

## Machine code

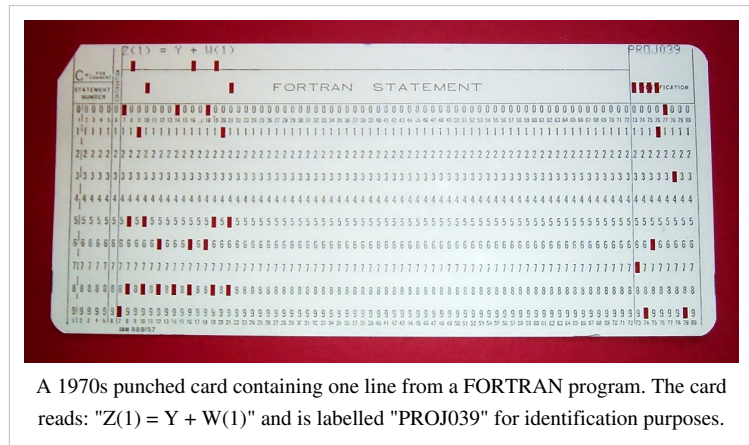
In most computers, individual instructions are stored as machine code with each instruction being given a unique number (its operation code or opcode for short). The command to add two numbers together would have one opcode, the command to multiply them would have a different opcode and so on. The simplest computers are able to perform any of a handful of different instructions; the more complex computers have several hundred to choose from, each with a unique numerical code. Since the computer's memory is able to store numbers, it can also store the instruction codes. This leads to the important fact that entire programs (which are just lists of these instructions) can be represented as lists of numbers and can themselves be manipulated inside the computer in the same way as numeric data. The fundamental concept of storing programs in the computer's memory alongside the data they operate on is the crux of the von Neumann, or stored program, architecture. In some cases, a computer might store some or all of its program in memory that is kept separate from the data it operates on. This is called the Harvard architecture after the Harvard Mark I computer. Modern von Neumann computers display some traits of the Harvard architecture in their designs, such as in CPU caches.

While it is possible to write computer programs as long lists of numbers (machine language) and while this technique was used with many early computers,<sup>[34]</sup> it is extremely tedious and potentially error-prone to do so in practice, especially for complicated programs. Instead, each basic instruction can be given a short name that is indicative of its function and easy to remember – a mnemonic such as ADD, SUB, MULT or JUMP. These mnemonics are collectively known as a computer's assembly language. Converting programs written in assembly language into something the computer can actually understand (machine language) is usually done by a computer program called an assembler.



## Programming language

Programming languages provide various ways of specifying programs for computers to run. Unlike natural languages, programming languages are designed to permit no ambiguity and to be concise. They are purely written languages and are often difficult to read aloud. They are generally either translated into machine code by a compiler or an assembler before being run, or translated directly at run time by an interpreter. Sometimes programs are executed by a hybrid method of the two techniques.



A 1970s punched card containing one line from a FORTRAN program. The card reads: "Z(1) = Y + W(1)" and is labelled "PROJ039" for identification purposes.

## Low-level languages

Machine languages and the assembly languages that represent them (collectively termed *low-level programming languages*) tend to be unique to a particular type of computer. For instance, an ARM architecture computer (such as may be found in a PDA or a hand-held videogame) cannot understand the machine language of an Intel Pentium or the AMD Athlon 64 computer that might be in a PC.<sup>[35]</sup>

## Higher-level languages

Though considerably easier than in machine language, writing long programs in assembly language is often difficult and is also error prone. Therefore, most practical programs are written in more abstract high-level programming languages that are able to express the needs of the programmer more conveniently (and thereby help reduce programmer error). High level languages are usually "compiled" into machine language (or sometimes into assembly language and then into machine language) using another computer program called a compiler.<sup>[36]</sup> High level languages are less related to the workings of the target computer than assembly language, and more related to the language and structure of the problem(s) to be solved by the final program. It is therefore often possible to use different compilers to translate the same high level language program into the machine language of many different types of computer. This is part of the means by which software like video games may be made available for different computer architectures such as personal computers and various video game consoles.

## Program design

Program design of small programs is relatively simple and involves the analysis of the problem, collection of inputs, using the programming constructs within languages, devising or using established procedures and algorithms, providing data for output devices and solutions to the problem as applicable. As problems become larger and more complex, features such as subprograms, modules, formal documentation, and new paradigms such as object-oriented programming are encountered. Large programs involving thousands of line of code and more require formal software methodologies. The task of developing large software systems presents a significant intellectual challenge. Producing software with an acceptably high reliability within a predictable schedule and budget has historically been difficult; the academic and professional discipline of software engineering concentrates specifically on this challenge.

## Components

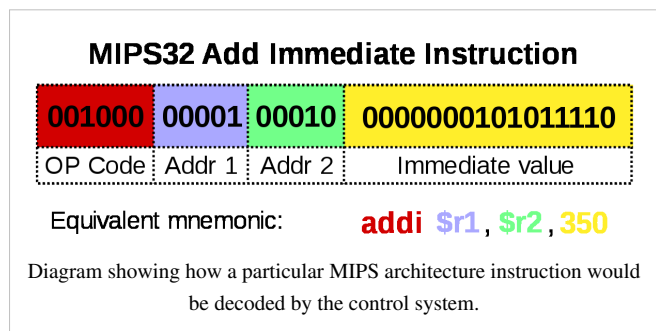
A general purpose computer has four main components: the arithmetic logic unit (ALU), the control unit, the memory, and the input and output devices (collectively termed I/O). These parts are interconnected by busses, often made of groups of wires.

Inside each of these parts are thousands to trillions of small electrical circuits which can be turned off or on by means of an electronic switch. Each circuit represents a bit (binary digit) of information so that when the circuit is on it represents a "1", and when off it represents a "0" (in positive logic representation). The circuits are arranged in logic gates so that one or more of the circuits may control the state of one or more of the other circuits.

The control unit, ALU, registers, and basic I/O (and often other hardware closely linked with these) are collectively known as a central processing unit (CPU). Early CPUs were composed of many separate components but since the mid-1970s CPUs have typically been constructed on a single integrated circuit called a *microprocessor*.

## Control unit

The control unit (often called a control system or central controller) manages the computer's various components; it reads and interprets (decodes) the program instructions, transforming them into a series of control signals which activate other parts of the computer.<sup>[37]</sup> Control systems in advanced computers may change the order of some instructions so as to improve performance.



A key component common to all CPUs is the program counter, a special memory cell (a register) that keeps track of which location in memory the next instruction is to be read from.<sup>[38]</sup>

The control system's function is as follows—note that this is a simplified description, and some of these steps may be performed concurrently or in a different order depending on the type of CPU:

1. Read the code for the next instruction from the cell indicated by the program counter.
2. Decode the numerical code for the instruction into a set of commands or signals for each of the other systems.
3. Increment the program counter so it points to the next instruction.
4. Read whatever data the instruction requires from cells in memory (or perhaps from an input device). The location of this required data is typically stored within the instruction code.
5. Provide the necessary data to an ALU or register.
6. If the instruction requires an ALU or specialized hardware to complete, instruct the hardware to perform the requested operation.
7. Write the result from the ALU back to a memory location or to a register or perhaps an output device.
8. Jump back to step (1).

Since the program counter is (conceptually) just another set of memory cells, it can be changed by calculations done in the ALU. Adding 100 to the program counter would cause the next instruction to be read from a place 100 locations further down the program. Instructions that modify the program counter are often known as "jumps" and allow for loops (instructions that are repeated by the computer) and often conditional instruction execution (both examples of control flow).

The sequence of operations that the control unit goes through to process an instruction is in itself like a short computer program, and indeed, in some more complex CPU designs, there is another yet smaller computer called a microsequencer, which runs a microcode program that causes all of these events to happen.

## Arithmetic logic unit (ALU)

The ALU is capable of performing two classes of operations: arithmetic and logic.<sup>[39]</sup>

The set of arithmetic operations that a particular ALU supports may be limited to addition and subtraction, or might include multiplication, division, trigonometry functions such as sine, cosine, etc., and square roots. Some can only operate on whole numbers (integers) whilst others use floating point to represent real numbers, albeit with limited precision. However, any computer that is capable of performing just the simplest operations can be programmed to break down the more complex operations into simple steps that it can perform. Therefore, any computer can be programmed to perform any arithmetic operation—although it will take more time to do so if its ALU does not directly support the operation. An ALU may also compare numbers and return boolean truth values (true or false) depending on whether one is equal to, greater than or less than the other ("is 64 greater than 65?").

Logic operations involve Boolean logic: AND, OR, XOR and NOT. These can be useful for creating complicated conditional statements and processing boolean logic.

Superscalar computers may contain multiple ALUs, allowing them to process several instructions simultaneously.<sup>[40]</sup> Graphics processors and computers with SIMD and MIMD features often contain ALUs that can perform arithmetic on vectors and matrices.

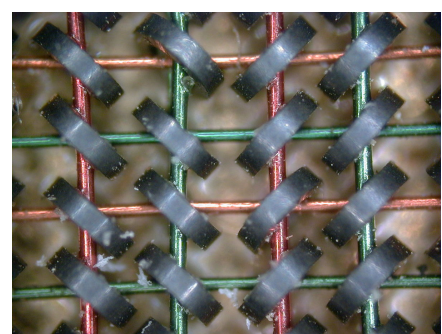
## Memory

A computer's memory can be viewed as a list of cells into which numbers can be placed or read. Each cell has a numbered "address" and can store a single number. The computer can be instructed to "put the number 123 into the cell numbered 1357" or to "add the number that is in cell 1357 to the number that is in cell 2468 and put the answer into cell 1595". The information stored in memory may represent practically anything. Letters, numbers, even computer instructions can be placed into memory with equal ease. Since the CPU does not differentiate between different types of information, it is the software's responsibility to give significance to what the memory sees as nothing but a series of numbers.

In almost all modern computers, each memory cell is set up to store binary numbers in groups of eight bits (called a byte). Each byte is able to represent 256 different numbers ( $2^8 = 256$ ); either from 0 to 255 or  $-128$  to  $+127$ . To store larger numbers, several consecutive bytes may be used (typically, two, four or eight). When negative numbers are required, they are usually stored in two's complement notation. Other arrangements are possible, but are usually not seen outside of specialized applications or historical contexts. A computer can store any kind of information in memory if it can be represented numerically. Modern computers have billions or even trillions of bytes of memory.

The CPU contains a special set of memory cells called registers that can be read and written to much more rapidly than the main memory area. There are typically between two and one hundred registers depending on the type of CPU. Registers are used for the most frequently needed data items to avoid having to access main memory every time data is needed. As data is constantly being worked on, reducing the need to access main memory (which is often slow compared to the ALU and control units) greatly increases the computer's speed.

Computer main memory comes in two principal varieties: random-access memory or RAM and read-only memory or ROM. RAM can be read and written to anytime the CPU commands it, but ROM is pre-loaded with data and software that never changes, therefore the CPU can only read from it. ROM is typically used to store the computer's initial start-up instructions. In general, the contents of RAM are erased when the power to the computer is turned off, but ROM retains its data indefinitely. In a PC, the ROM contains a specialized program called the BIOS that



Magnetic core memory was the computer memory of choice throughout the 1960s, until it was replaced by semiconductor memory.



orchestrates loading the computer's operating system from the hard disk drive into RAM whenever the computer is turned on or reset. In embedded computers, which frequently do not have disk drives, all of the required software may be stored in ROM. Software stored in ROM is often called firmware, because it is notionally more like hardware than software. Flash memory blurs the distinction between ROM and RAM, as it retains its data when turned off but is also rewritable. It is typically much slower than conventional ROM and RAM however, so its use is restricted to applications where high speed is unnecessary.<sup>[41]</sup>

In more sophisticated computers there may be one or more RAM cache memories, which are slower than registers but faster than main memory. Generally computers with this sort of cache are designed to move frequently needed data into the cache automatically, often without the need for any intervention on the programmer's part.

## Input/output (I/O)

I/O is the means by which a computer exchanges information with the outside world.<sup>[42]</sup> Devices that provide input or output to the computer are called peripherals.<sup>[43]</sup> On a typical personal computer, peripherals include input devices like the keyboard and mouse, and output devices such as the display and printer. Hard disk drives, floppy disk drives and optical disc drives serve as both input and output devices. Computer networking is another form of I/O.

I/O devices are often complex computers in their own right, with their own CPU and memory. A graphics processing unit might contain fifty or more tiny computers that perform the calculations necessary to display 3D graphics. Modern desktop computers contain many smaller computers that assist the main CPU in performing I/O.



Hard disk drives are common storage devices used with computers.

## Multitasking

While a computer may be viewed as running one gigantic program stored in its main memory, in some systems it is necessary to give the appearance of running several programs simultaneously. This is achieved by multitasking i.e. having the computer switch rapidly between running each program in turn.<sup>[44]</sup>

One means by which this is done is with a special signal called an interrupt, which can periodically cause the computer to stop executing instructions where it was and do something else instead. By remembering where it was executing prior to the interrupt, the computer can return to that task later. If several programs are running "at the same time", then the interrupt generator might be causing several hundred interrupts per second, causing a program switch each time. Since modern computers typically execute instructions several orders of magnitude faster than human perception, it may appear that many programs are running at the same time even though only one is ever executing in any given instant. This method of multitasking is sometimes termed "time-sharing" since each program is allocated a "slice" of time in turn.<sup>[45]</sup>

Before the era of cheap computers, the principal use for multitasking was to allow many people to share the same computer.

Seemingly, multitasking would cause a computer that is switching between several programs to run more slowly, in direct proportion to the number of programs it is running, but most programs spend much of their time waiting for slow input/output devices to complete their tasks. If a program is waiting for the user to click on the mouse or press a key on the keyboard, then it will not take a "time slice" until the event it is waiting for has occurred. This frees up time for other programs to execute so that many programs may be run simultaneously without unacceptable speed loss.

## Multiprocessing

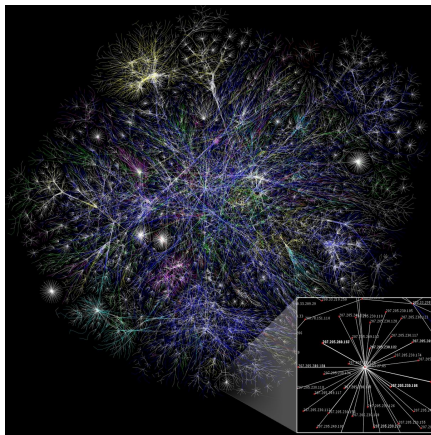
Some computers are designed to distribute their work across several CPUs in a multiprocessing configuration, a technique once employed only in large and powerful machines such as supercomputers, mainframe computers and servers. Multiprocessor and multi-core (multiple CPUs on a single integrated circuit) personal and laptop computers are now widely available, and are being increasingly used in lower-end markets as a result.

Supercomputers in particular often have highly unique architectures that differ significantly from the basic stored-program architecture and from general purpose computers.<sup>[46]</sup> They often feature thousands of CPUs, customized high-speed interconnects, and specialized computing hardware. Such designs tend to be useful only for specialized tasks due to the large scale of program organization required to successfully utilize most of the available resources at once. Supercomputers usually see usage in large-scale simulation, graphics rendering, and cryptography applications, as well as with other so-called "embarrassingly parallel" tasks.



Cray designed many supercomputers that used multiprocessing heavily.

## Networking and the Internet



Visualization of a portion of the routes on the Internet.

Computers have been used to coordinate information between multiple locations since the 1950s. The U.S. military's SAGE system was the first large-scale example of such a system, which led to a number of special-purpose commercial systems such as Sabre.<sup>[47]</sup>

In the 1970s, computer engineers at research institutions throughout the United States began to link their computers together using telecommunications technology. The effort was funded by ARPA (now DARPA), and the computer network that resulted was called the ARPANET.<sup>[48]</sup> The technologies that made the Arpanet possible spread and evolved.

In time, the network spread beyond academic and military institutions and became known as the Internet. The emergence of networking involved a redefinition of the nature and boundaries of the computer.

Computer operating systems and applications were modified to include the ability to define and access the resources of other computers on the network, such as peripheral devices, stored information, and the like, as extensions of the resources of an individual computer. Initially these facilities were available primarily to people working in high-tech environments, but in the 1990s the spread of applications like e-mail and the World Wide Web, combined with the development of cheap, fast networking technologies like Ethernet and ADSL saw computer networking become almost ubiquitous. In fact, the number of computers that are networked is growing phenomenally. A very large proportion of personal computers regularly connect to the Internet to communicate and receive information. "Wireless" networking, often utilizing mobile phone networks, has meant networking is becoming increasingly ubiquitous even in mobile computing environments.



## Computer architecture paradigms

There are many types of computer architectures:

- Quantum computer vs Chemical computer
- Scalar processor vs Vector processor
- Non-Uniform Memory Access (NUMA) computers
- Register machine vs Stack machine
- Harvard architecture vs von Neumann architecture
- Cellular architecture

The quantum computer architecture holds the most promise to revolutionize computing.<sup>[49]</sup>

Logic gates are a common abstraction which can apply to most of the above digital or analog paradigms.

The ability to store and execute lists of instructions called programs makes computers extremely versatile, distinguishing them from calculators. The Church–Turing thesis is a mathematical statement of this versatility: any computer with a minimum capability (being Turing-complete) is, in principle, capable of performing the same tasks that any other computer can perform. Therefore any type of computer (netbook, supercomputer, cellular automaton, etc.) is able to perform the same computational tasks, given enough time and storage capacity.

## Misconceptions

A computer does not need to be electronic, nor even have a processor, nor RAM, nor even a hard disk. While popular usage of the word "computer" is synonymous with a personal electronic computer, the modern<sup>[50]</sup> definition of a computer is literally "*A device that computes*, especially a programmable [usually] electronic machine that performs high-speed mathematical or logical operations or that assembles, stores, correlates, or otherwise processes information".<sup>[51]</sup> Any device which *processes information* qualifies as a computer, especially if the processing is purposeful.

## Required technology

Historically, computers evolved from mechanical computers and eventually from vacuum tubes to transistors. However, conceptually computational systems as flexible as a personal computer can be built out of almost anything. For example, a computer can be made out of billiard balls (billiard ball computer); an oft-quoted example. More realistically, modern computers are made out of transistors made of photolithographed semiconductors.

There is active research to make computers out of many promising new types of technology, such as optical computers, DNA computers, neural computers, and quantum computers. Most computers are universal, and are able to calculate any computable function, and are limited only by their memory capacity and operating speed. However different designs of computers can give very different performance for particular problems; for example quantum computers can potentially break some modern encryption algorithms (by quantum factoring) very quickly.

## Further topics

- Glossary of computers

## Artificial intelligence

A computer will solve problems in exactly the way it is programmed to, without regard to efficiency, alternative solutions, possible shortcuts, or possible errors in the code. Computer programs that learn and adapt are part of the emerging field of artificial intelligence and machine learning.

## Hardware

The term *hardware* covers all of those parts of a computer that are tangible objects. Circuits, displays, power supplies, cables, keyboards, printers and mice are all hardware.

## History of computing hardware

First generation (mechanical/electromechanical)	Calculators	Antikythera mechanism, Difference engine, Norden bombsight
	Programmable devices	Jacquard loom, Analytical engine, Harvard Mark I, Z3
Second generation (vacuum tubes)	Calculators	Atanasoff–Berry Computer, IBM 604, UNIVAC 60, UNIVAC 120
	Programmable devices	Colossus, ENIAC, Manchester Small-Scale Experimental Machine, EDSAC, Manchester Mark 1, Ferranti Pegasus, Ferranti Mercury, CSIRAC, EDVAC, UNIVAC I, IBM 701, IBM 702, IBM 650, Z22
Third generation (discrete transistors and SSI, MSI, LSI integrated circuits)	Mainframes	IBM 7090, IBM 7080, IBM System/360, BUNCH
	Minicomputer	PDP-8, PDP-11, IBM System/32, IBM System/36
Fourth generation (VLSI integrated circuits)	Minicomputer	VAX, IBM System i
	4-bit microcomputer	Intel 4004, Intel 4040
	8-bit microcomputer	Intel 8008, Intel 8080, Motorola 6800, Motorola 6809, MOS Technology 6502, Zilog Z80
	16-bit microcomputer	Intel 8088, Zilog Z8000, WDC 65816/65802
	32-bit microcomputer	Intel 80386, Pentium, Motorola 68000, ARM architecture
	64-bit microcomputer <sup>[52]</sup>	Alpha, MIPS, PA-RISC, PowerPC, SPARC, x86-64
	Embedded computer	Intel 8048, Intel 8051
	Personal computer	Desktop computer, Home computer, Laptop computer, Personal digital assistant (PDA), Portable computer, Tablet PC, Wearable computer
Theoretical/experimental	Quantum computer, Chemical computer, DNA computing, Optical computer, Spintronics based computer	

## Other hardware topics

Peripheral device (input/output)	Input	Mouse, keyboard, joystick, image scanner, webcam, graphics tablet, microphone
	Output	Monitor, printer, loudspeaker
	Both	Floppy disk drive, hard disk drive, optical disc drive, teleprinter
Computer busses	Short range	RS-232, SCSI, PCI, USB
	Long range (computer networking)	Ethernet, ATM, FDDI

## Software

*Software* refers to parts of the computer which do not have a material form, such as programs, data, protocols, etc. When software is stored in hardware that cannot easily be modified (such as BIOS ROM in an IBM PC compatible), it is sometimes called "firmware".

Operating system	Unix and BSD	UNIX System V, IBM AIX, HP-UX, Solaris (SunOS), IRIX, List of BSD operating systems
	GNU/Linux	List of Linux distributions, Comparison of Linux distributions
	Microsoft Windows	Windows 95, Windows 98, Windows NT, Windows 2000, Windows Me, Windows XP, Windows Vista, Windows 7, Windows 8
	DOS	86-DOS (QDOS), PC-DOS, MS-DOS, DR-DOS, FreeDOS
	Mac OS	Mac OS classic, Mac OS X
	Embedded and real-time	List of embedded operating systems
	Experimental	Amoeba, Oberon/Bluebottle, Plan 9 from Bell Labs
Library	Multimedia	DirectX, OpenGL, OpenAL
	Programming library	C standard library, Standard Template Library
Data	Protocol	TCP/IP, Kermit, FTP, HTTP, SMTP
	File format	HTML, XML, JPEG, MPEG, PNG
User interface	Graphical user interface (WIMP)	Microsoft Windows, GNOME, KDE, QNX Photon, CDE, GEM, Aqua
	Text-based user interface	Command-line interface, Text user interface

Application	Office suite	Word processing, Desktop publishing, Presentation program, Database management system, Scheduling & Time management, Spreadsheet, Accounting software
	Internet Access	Browser, E-mail client, Web server, Mail transfer agent, Instant messaging
	Design and manufacturing	Computer-aided design, Computer-aided manufacturing, Plant management, Robotic manufacturing, Supply chain management
	Graphics	Raster graphics editor, Vector graphics editor, 3D modeler, Animation editor, 3D computer graphics, Video editing, Image processing
	Audio	Digital audio editor, Audio playback, Mixing, Audio synthesis, Computer music
	Software engineering	Compiler, Assembler, Interpreter, Debugger, Text editor, Integrated development environment, Software performance analysis, Revision control, Software configuration management
	Educational	Edutainment, Educational game, Serious game, Flight simulator
	Games	Strategy, Arcade, Puzzle, Simulation, First-person shooter, Platform, Massively multiplayer, Interactive fiction
	Misc	Artificial intelligence, Antivirus software, Malware scanner, Installer/Package management systems, File manager

## Languages

There are thousands of different programming languages—some intended to be general purpose, others useful only for highly specialized applications.

### Programming languages

Lists of programming languages	Timeline of programming languages, List of programming languages by category, Generational list of programming languages, List of programming languages, Non-English-based programming languages
Commonly used assembly languages	ARM, MIPS, x86
Commonly used high-level programming languages	Ada, BASIC, C, C++, C#, COBOL, Fortran, Java, Lisp, Pascal, Object Pascal
Commonly used scripting languages	Bourne script, JavaScript, Python, Ruby, PHP, Perl

## Professions and organizations

As the use of computers has spread throughout society, there are an increasing number of careers involving computers.

### Computer-related professions

Hardware-related	Electrical engineering, Electronic engineering, Computer engineering, Telecommunications engineering, Optical engineering, Nanoengineering
Software-related	Computer science, Computer engineering, Desktop publishing, Human–computer interaction, Information technology, Information systems, Computational science, Software engineering, Video game industry, Web design

The need for computers to work well together and to be able to exchange information has spawned the need for many standards organizations, clubs and societies of both a formal and informal nature.

## Organizations

Standards groups	ANSI, IEC, IEEE, IETF, ISO, W3C
Professional Societies	ACM, AIS, IET, IFIP, BCS
Free/Open source software groups	Free Software Foundation, Mozilla Foundation, Apache Software Foundation

## Notes

- [1] In 1946, ENIAC required an estimated 174 kW. By comparison, a modern laptop computer may use around 30 W; nearly six thousand times less. "Approximate Desktop & Notebook Power Usage" (<http://www.upenn.edu/computing/provider/docs/hardware/powerusage.html>). University of Pennsylvania. . Retrieved 20 June 2009.
- [2] Early computers such as Colossus and ENIAC were able to process between 5 and 100 operations per second. A modern "commodity" microprocessor (as of 2007) can process billions of operations per second, and many of these operations are more complicated and useful than early computer operations. "Intel Core2 Duo Mobile Processor: Features" ([http://www.intel.com/cd/channel/reseller/asm-na/eng/products/mobile/processors/core2duo\\_m/feature/index.htm](http://www.intel.com/cd/channel/reseller/asm-na/eng/products/mobile/processors/core2duo_m/feature/index.htm)). Intel Corporation. . Retrieved 20 June 2009.
- [3] *computer*, n. (<http://dictionary.oed.com/>). Oxford English Dictionary (2 ed.). Oxford University Press. 1989. . Retrieved 10 April 2009.
- [4] \* Ifrah, Georges (2001). *The Universal History of Computing: From the Abacus to the Quantum Computer*. New York: John Wiley & Sons. ISBN 0-471-39671-0. From 2700 to 2300 BC, Georges Ifrah, pp.11
- [5] Berkeley, Edmund (1949). *Giant Brains, or Machines That Think*. John Wiley & Sons. p. 19. Edmund Berkeley
- [6] According to advertising on Pickett's N600 slide rule boxes. "Pickett Apollo Box Scans" (<http://copland.udel.edu/~mm/sliderule/lem/>). Copland.udel.edu. . Retrieved 20 February 2010.
- [7] "Discovering How Greeks Computed in 100 B.C." (<http://www.nytimes.com/2008/07/31/science/31computer.html?hp>). *The New York Times*. 31 July 2008. . Retrieved 27 March 2010.
- [8] "Heron of Alexandria" (<http://www.mlahanas.de/Greeks/HeronAlexandria2.htm>). . Retrieved 15 January 2008.
- [9] Felt, Dorr E. (1916). *Mechanical arithmetic, or The history of the counting machine* (<http://www.archive.org/details/mechanicalarithm00feltrich>). Chicago: Washington Institute. p. 8. . Dorr E. Felt
- [10] "Speaking machines" ([http://books.google.co.uk/books?id=Xt4PAAAAAYAAJ&pg=PT38&dq=the+parlour+review+january+1838&hl=en&ei=0yqzTN3kLMTHswa2wMjSDQ&sa=X&oi=book\\_result&ct=result&resnum=1&ved=0CCsQ6AEwAA#v=onepage&q&f=false](http://books.google.co.uk/books?id=Xt4PAAAAAYAAJ&pg=PT38&dq=the+parlour+review+january+1838&hl=en&ei=0yqzTN3kLMTHswa2wMjSDQ&sa=X&oi=book_result&ct=result&resnum=1&ved=0CCsQ6AEwAA#v=onepage&q&f=false)). *The parlour review, Philadelphia* 1 (3). 20 January 1838. . Retrieved 11 October 2010.
- [11] "Pascal and Leibnitz, in the seventeenth century, and Diderot at a later period, endeavored to construct a machine which might serve as a substitute for human intelligence in the combination of figures" *The Gentleman's magazine*, Volume 202, p.100 ([http://books.google.fr/books?id=Rf0IAAAAIAAJ&pg=PA100&dq=arithmometer&as\\_brr=1#v=onepage&q=arithmometer&f=false](http://books.google.fr/books?id=Rf0IAAAAIAAJ&pg=PA100&dq=arithmometer&as_brr=1#v=onepage&q=arithmometer&f=false))
- [12] Babbage's Difference engine in 1823 and his Analytical engine in the mid-1830s
- [13] "It is reasonable to inquire, therefore, whether it is possible to devise a machine which will do for mathematical computation what the automatic lathe has done for engineering. The first suggestion that such a machine could be made came more than a hundred years ago from the mathematician Charles Babbage. Babbage's ideas have only been properly appreciated in the last ten years, but we now realize that he understood clearly all the fundamental principles which are embodied in modern digital computers" *Faster than thought*, edited by B. V. Bowden, 1953, Pitman publishing corporation
- [14] "...Among this extraordinary galaxy of talent Charles Babbage appears to be one of the most remarkable of all. Most of his life he spent in an entirely unsuccessful attempt to make a machine which was regarded by his contemporaries as utterly preposterous, and his efforts were regarded as futile, time-consuming and absurd. In the last decade or so we have learnt how his ideas can be embodied in a modern digital computer. He understood more about the logic of these machines than anyone else in the world had learned until after the end of the last war" Foreword, *Irascible Genius, Charles Babbage, inventor* by Mabothe Moseley, 1964, London, Hutchinson
- [15] In the proposal that Aiken gave IBM in 1937 while requesting funding for the Harvard Mark I we can read: "Few **calculating machines** have been designed strictly for application to scientific investigations, the notable exceptions being those of Charles Babbage and others who followed him ... After abandoning the difference engine, Babbage devoted his energy to the design and construction of an analytical engine of far higher powers than the difference engine ... Since the time of Babbage, the development of **calculating machinery** has continued at an increasing rate." Howard Aiken, *Proposed automatic calculating machine*, reprinted in: *The origins of Digital computers, Selected Papers*, Edited by Brian Randell, 1973, ISBN 3-540-06169-X
- [16] "Intel Museum – The 4004, Big deal then, Big deal now" (<http://www.intel.com/about/companyinfo/museum/exhibits/4004/index.htm>). Intel.com. . Retrieved 29 January 2012.
- [17] From cave paintings to the internet (<http://www.historyofscience.com/G2I/timeline/index.php?category=Computers+&+the+Human+Brain>) HistoryofScience.com
- [18] See: Anthony Hyman, ed., *Science and Reform: Selected Works of Charles Babbage* (Cambridge, England: Cambridge University Press, 1989), page 298. It is in the collection of the Science Museum in London, England. (Delve (2007), page 99.)
- [19] The analytical engine should not be confused with Babbage's difference engine which was a non-programmable mechanical calculator.

- [20] "Columbia University Computing History: Herman Hollerith" (<http://www.columbia.edu/acis/history/hollerith.html>). Columbia.edu. . Retrieved 11 December 2010.
- [21] "Alan Turing – Time 100 People of the Century" (<http://205.188.238.181/time/time100/scientist/profile/turing.html>). Time Magazine. . Retrieved 13 June 2009. "The fact remains that everyone who taps at a keyboard, opening a spreadsheet or a word-processing program, is working on an incarnation of a Turing machine"
- [22] "John Vincent Atanasoff and the Birth of Electronic Digital Computing" (<http://www.cs.iastate.edu/jva/jva-archive.shtml>). Cs.iastate.edu. . Retrieved 29 January 2012.
- [23] "John Vincent Atanasoff – the father of the computer" (<http://www.columbia.edu/~td2177/JVAtanasoff/JVAtanasoff.html>). Columbia.edu. . Retrieved 29 January 2012.
- [24] "Atanasoff-Berry Computer" (<http://energysciencenews.com/phpBB3/viewtopic.php?f=1&t=98&p=264#p264>). . Retrieved 20 November 2010.
- [25] "Spiegel: The inventor of the computer's biography was published" (<http://www.spiegel.de/netzwelt/gadgets/0,1518,651776,00.html>). *Der Spiegel*. 28 September 2009. . Retrieved 11 December 2010.
- [26] "Inventor Profile: George R. Stibitz" ([http://www.invent.org/hall\\_of\\_fame/140.html](http://www.invent.org/hall_of_fame/140.html)). National Inventors Hall of Fame Foundation, Inc.. .
- [27] Rojas, R. (1998). "How to make Zuse's Z3 a universal computer". *IEEE Annals of the History of Computing* **20** (3): 51–54. doi:10.1109/85.707574.
- [28] B. Jack Copeland, ed., *Colossus: The Secrets of Bletchley Park's Codebreaking Computers*, Oxford University Press, 2006
- [29] "Robot Mathematician Knows All The Answers", *October 1944, Popular Science* ([http://books.google.com/books?id=PyEDAAAAMBAJ&pg=PA86&dq=motor+gun+boat&hl=en&ei=LxTqTMfGI4-bnwfEyNiWDQ&sa=X&oi=book\\_result&ct=result&resnum=6&ved=0CEIQ6AEwBQ#v=onepage&q=motor+gun+boat&f=true](http://books.google.com/books?id=PyEDAAAAMBAJ&pg=PA86&dq=motor+gun+boat&hl=en&ei=LxTqTMfGI4-bnwfEyNiWDQ&sa=X&oi=book_result&ct=result&resnum=6&ved=0CEIQ6AEwBQ#v=onepage&q=motor+gun+boat&f=true)). Google Books. . Retrieved 11 December 2010.
- [30] Lavington 1998, p. 37
- [31] This program was written similarly to those for the PDP-11 minicomputer and shows some typical things a computer can do. All the text after the semicolons are comments for the benefit of human readers. These have no significance to the computer and are ignored. (Digital Equipment Corporation 1972)
- [32] It is not universally true that bugs are solely due to programmer oversight. Computer hardware may fail or may itself have a fundamental problem that produces unexpected results in certain situations. For instance, the Pentium FDIV bug caused some Intel microprocessors in the early 1990s to produce inaccurate results for certain floating point division operations. This was caused by a flaw in the microprocessor design and resulted in a partial recall of the affected devices.
- [33] Taylor, Alexander L., III (16 April 1984). "The Wizard Inside the Machine" (<http://www.time.com/time/printout/0,8816,954266,00.html>). *TIME*. . Retrieved 17 February 2007.
- [34] Even some later computers were commonly programmed directly in machine code. Some minicomputers like the DEC PDP-8 could be programmed directly from a panel of switches. However, this method was usually used only as part of the booting process. Most modern computers boot entirely automatically by reading a boot program from some non-volatile memory.
- [35] However, there is sometimes some form of machine language compatibility between different computers. An x86-64 compatible microprocessor like the AMD Athlon 64 is able to run most of the same programs that an Intel Core 2 microprocessor can, as well as programs designed for earlier microprocessors like the Intel Pentiums and Intel 80486. This contrasts with very early commercial computers, which were often one-of-a-kind and totally incompatible with other computers.
- [36] High level languages are also often interpreted rather than compiled. Interpreted languages are translated into machine code on the fly, while running, by another program called an interpreter.
- [37] The control unit's role in interpreting instructions has varied somewhat in the past. Although the control unit is solely responsible for instruction interpretation in most modern computers, this is not always the case. Many computers include some instructions that may only be partially interpreted by the control system and partially interpreted by another device. This is especially the case with specialized computing hardware that may be partially self-contained. For example, EDVAC, one of the earliest stored-program computers, used a central control unit that only interpreted four instructions. All of the arithmetic-related instructions were passed on to its arithmetic unit and further decoded there.
- [38] Instructions often occupy more than one memory address, therefore the program counter usually increases by the number of memory locations required to store one instruction.
- [39] David J. Eck (2000). *The Most Complex Machine: A Survey of Computers and Computing*. A K Peters, Ltd.. p. 54. ISBN 978-1-56881-128-4.
- [40] Erricos John Kontoghiorghes (2006). *Handbook of Parallel Computing and Statistics*. CRC Press. p. 45. ISBN 978-0-8247-4067-2.
- [41] Flash memory also may only be rewritten a limited number of times before wearing out, making it less useful for heavy random access usage. (Verma & Mielke 1988)
- [42] Donald Eadie (1968). *Introduction to the Basic Computer*. Prentice-Hall. p. 12.
- [43] Arpad Barna; Dan I. Porat (1976). *Introduction to Microcomputers and the Microprocessors*. Wiley. p. 85. ISBN 978-0-471-05051-3.
- [44] Jerry Peek; Grace Todino, John Strang (2002). *Learning the UNIX Operating System: A Concise Guide for the New User*. O'Reilly. p. 130. ISBN 978-0-596-00261-9.
- [45] Gillian M. Davis (2002). *Noise Reduction in Speech Applications*. CRC Press. p. 111. ISBN 978-0-8493-0949-6.
- [46] However, it is also very common to construct supercomputers out of many pieces of cheap commodity hardware; usually individual computers connected by networks. These so-called computer clusters]] can often provide supercomputer performance at a much lower cost

- than customized designs. While custom architectures are still used for most of the most powerful supercomputers, there has been a proliferation of cluster computers in recent years. (TOP500 2006)
- [47] Agatha C. Hughes (2000). *Systems, Experts, and Computers*. MIT Press. p. 161. ISBN 978-0-262-08285-3. "The experience of SAGE helped make possible the first truly large-scale commercial real-time network: the SABRE computerized airline reservations system..."
- [48] "A Brief History of the Internet" (<http://www.isoc.org/internet/history/brief.shtml>). Internet Society. . Retrieved 20 September 2008.
- [49] "Computer architecture: fundamentals and principles of computer design" ([http://books.google.com/books?id=ZWaUurOwMPQC&q=quantum+computers&dq=insufficient+address+computer+architecture&source=gbs\\_word\\_cloud\\_r&cad=3#v=snippet&q=quantum+computers&f=false](http://books.google.com/books?id=ZWaUurOwMPQC&q=quantum+computers&dq=insufficient+address+computer+architecture&source=gbs_word_cloud_r&cad=3#v=snippet&q=quantum+computers&f=false)) by Joseph D. Dumas 2006. page 340.
- [50] According to the *Shorter Oxford English Dictionary* (6th ed, 2007), the word *computer* dates back to the mid 17th century, when it referred to "A person who makes calculations; specifically a person employed for this in an observatory etc."
- [51] "Definition of computer" (<http://thefreedictionary.com/computer>). Thefreedictionary.com. . Retrieved 29 January 2012.
- [52] Most major 64-bit instruction set architectures are extensions of earlier designs. All of the architectures listed in this table, except for Alpha, existed in 32-bit forms before their 64-bit incarnations were introduced.

## References

- <sup>a</sup> Kempf, Karl (1961). *Historical Monograph: Electronic Computers Within the Ordnance Corps* (<http://ed-thelen.org/comp-hist/U-S-Ord-61.html>). Aberdeen Proving Ground (United States Army).
- <sup>a</sup> Phillips, Tony (2000). "The Antikythera Mechanism I" (<http://www.math.sunysb.edu/~tony/whatsnew/column/antikytheraI-0400/kyth1.html>). American Mathematical Society. Retrieved 5 April 2006.
- <sup>a</sup> Shannon, Claude Elwood (1940). *A symbolic analysis of relay and switching circuits* (<http://hdl.handle.net/1721.1/11173>). Massachusetts Institute of Technology.
- Digital Equipment Corporation (1972) (PDF). *PDP-11/40 Processor Handbook* ([http://bitsavers.vt100.net/dec/www.computer.museum.uq.edu.au\\_mirror/D-09-30\\_PDP11-40\\_Processor\\_Handbook.pdf](http://bitsavers.vt100.net/dec/www.computer.museum.uq.edu.au_mirror/D-09-30_PDP11-40_Processor_Handbook.pdf)). Maynard, MA: Digital Equipment Corporation.
- Verma, G.; Mielke, N. (1988). *Reliability performance of ETOX based flash memories*. IEEE International Reliability Physics Symposium.
- Meuer, Hans; Strohmaier, Erich; Simon, Horst; Dongarra, Jack (13 November 2006). "Architectures Share Over Time" (<http://www.top500.org/lists/2006/11/overtime/Architectures>). TOP500. Retrieved 27 November 2006.
- Lavington, Simon (1998). *A History of Manchester Computers* (2 ed.). Swindon: The British Computer Society. ISBN 978-0-902505-01-8.
- Stokes, Jon (2007). *Inside the Machine: An Illustrated Introduction to Microprocessors and Computer Architecture*. San Francisco: No Starch Press. ISBN 978-1-59327-104-6.
- Felt, Dorr E. (1916). *Mechanical arithmetic, or The history of the counting machine* (<http://www.archive.org/details/mechanicalarithm00feltrich>). Chicago: Washington Institute.
- Ifrah, Georges (2001). *The Universal History of Computing: From the Abacus to the Quantum Computer*. New York: John Wiley & Sons. ISBN 0-471-39671-0.
- Berkeley, Edmund (1949). *Giant Brains, or Machines That Think*. John Wiley & Sons.

## External links

- A Brief History of Computing (<http://www.life.com/image/first/in-gallery/48681/click-a-brief-history-of-computing#index/0>) – slideshow by *Life magazine*



# Antenna (radio)

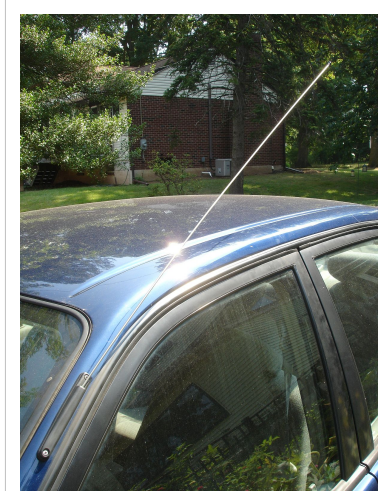
An **antenna** (or **aerial**) is an electrical device which converts electric power into radio waves, and vice versa. It is usually used with a radio transmitter or radio receiver. In transmission, a radio transmitter supplies an oscillating radio frequency electric current to the antenna's terminals, and the antenna radiates the energy from the current as electromagnetic waves (radio waves). In reception, an antenna intercepts some of the power of an electromagnetic wave in order to produce a tiny voltage at its terminals, that is applied to a receiver to be amplified.

Antennas are essential components of all equipment that uses radio. They are used in systems such as radio broadcasting, broadcast television, two-way radio, communications receivers, radar, cell phones, and satellite communications, as well as other devices such as garage door openers, wireless microphones, bluetooth enabled devices, wireless computer networks, baby monitors, and RFID tags on merchandise.

Typically an antenna consists of an arrangement of metallic conductors ("elements"), electrically connected (often through a transmission line) to the receiver or transmitter. An oscillating current of electrons forced through the antenna by a transmitter will create an oscillating magnetic field around the antenna elements, while the charge of the electrons also creates an oscillating electric field along the elements. These time-varying fields, when created in the proper proportions, radiate away from the antenna into space as a moving transverse electromagnetic field wave. Conversely, during reception, the oscillating electric and magnetic fields of an incoming radio wave exert force on the electrons in the antenna elements, causing them to move back and forth, creating oscillating currents in the antenna.

Antennas may also contain reflective or directive elements or surfaces not connected to the transmitter or receiver, such as parasitic elements, parabolic reflectors or horns, which serve to direct the radio waves into a beam or other desired radiation pattern. Antennas can be designed to transmit or receive radio waves in all directions equally (omnidirectional antennas), or transmit them in a beam in a particular direction, and receive from that one direction only (directional or high gain antennas).

The first antennas were built in 1888 by German physicist Heinrich Hertz in his pioneering experiments to prove the existence of electromagnetic waves predicted by the theory of James Clerk Maxwell. Hertz placed dipole antennas at the focal point of parabolic reflectors for both transmitting and receiving. He published his work in *Annalen der Physik und Chemie* (vol. 36, 1889).



Whip antenna on car

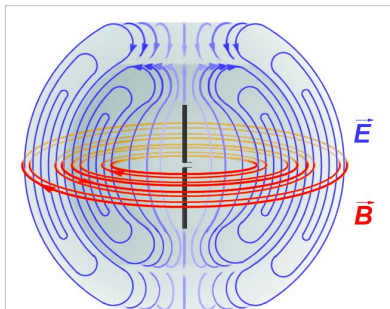


Diagram of the electric fields (*blue*) and magnetic fields (*red*) radiated by a dipole antenna (*black rods*) during transmission.



Large parabolic antenna for communicating with spacecraft

### Terminology

The words *antenna* (plural: *antennas*<sup>[1]</sup>) and *aerial* are used interchangeably. Occasionally a rigid metallic structure is called an "antenna" while the wire form is called an "aerial". However, note the important international technical journal, the *IEEE Transactions on Antennas and Propagation*.<sup>[2]</sup> In the United Kingdom and other areas where British English is used, the term aerial is sometimes used although 'antenna' has been universal in professional use for many years.

The origin of the word *antenna* relative to wireless apparatus is attributed to Italian radio pioneer Guglielmo Marconi. In 1895, while testing early radio apparatus in the Swiss Alps at Salvan, Switzerland in the Mont Blanc region, Marconi experimented with long wire "aerials". He used a 2.5 meter vertical pole, with a wire attached to the top running down to the transmitter, as a radiating and receiving aerial element. In Italian a tent pole is known as *l'antenna centrale*, and the pole with the wire was simply called *l'antenna*. Until then wireless radiating transmitting and receiving elements were known simply as aeriels or terminals. Because of his prominence, Marconi's use of the word *antenna* (Italian for *pole*) spread among wireless researchers, and later to the general public.<sup>[3]</sup>

In common usage, the word *antenna* may refer broadly to an entire assembly including support structure, enclosure (if any), etc. in addition to the actual functional components. Especially at microwave frequencies, a receiving antenna may include not only the actual electrical antenna but an integrated preamplifier or mixer.



Rooftop television antennas in Israel. Yagi-Uda antennas like these six are widely used at VHF and UHF frequencies.



"Rabbit ears" dipole antenna for television reception



Cell phone base station antennas



Wi-Fi WestNet Wi-Fi base station antennas in Calgary, Alberta



Parabolic antenna by Himalaya Television Nepal



Yagi antenna used for mobile military communications station, Dresden, Germany, 1955



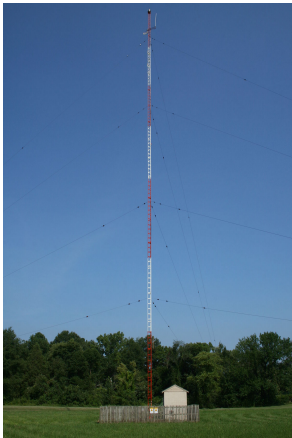
Turnstile type transmitting antenna for VHF low band television broadcasting station, Germany.



Folded dipole antenna



Large Yagi antenna used by amateur radio hobbyists



A mast radiator antenna for an AM radio station in Chapel Hill, North Carolina

## Overview

Antennas are required by any radio receiver or transmitter to couple its electrical connection to the electromagnetic field. Radio waves are electromagnetic waves which carry signals through the air (or through space) at the speed of light with almost no transmission loss. Radio transmitters and receivers are used to convey signals (information) in systems including broadcast (audio) radio, television, mobile telephones, wi-fi (WLAN) data networks, trunk lines and point-to-point communications links (telephone, data networks), satellite links, many remote controlled devices such as garage door openers, and wireless remote sensors, among many others. Radio waves are also used directly for measurements in technologies including RADAR, GPS, and radio astronomy. In each and every case, the transmitters and receivers involved require antennas, although these are sometimes hidden (such as the antenna inside an AM radio or inside a laptop computer equipped with wi-fi).



Antennas of the Atacama Large Millimeter submillimeter Array.<sup>[4]</sup>

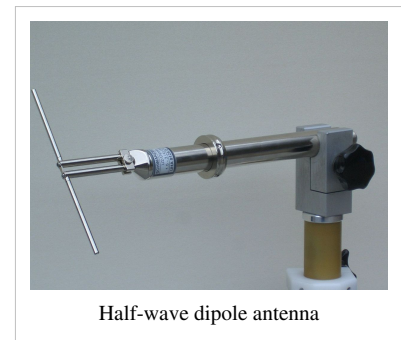
According to their applications and technology available, antennas generally fall in one of two categories:

1. Omnidirectional or only weakly directional antennas which receive or radiate more or less in all directions. These are employed when the relative position of the other station is unknown or arbitrary. They are also used at lower frequencies where a directional antenna would be too large, or simply to cut costs in applications where a directional antenna isn't required.
2. Directional or *beam* antennas which are intended to preferentially radiate or receive in a particular direction or directional pattern.

In common usage "omnidirectional" usually refers to all horizontal directions, typically with reduced performance in the direction of the sky or the ground (a truly isotropic radiator is not even possible). A "directional" antenna usually is intended to maximize its coupling to the electromagnetic field in the direction of the other station, or sometimes to cover a particular sector such as a 120° horizontal fan pattern in the case of a panel antenna at a cell site.

One example of omnidirectional antennas is the very common *vertical antenna* or whip antenna consisting of a metal rod (often, but not always, a quarter of a wavelength long). A dipole antenna is similar but consists of two such conductors extending in opposite directions, with a total length that is often, but not always, a half of a wavelength long. Dipoles are typically oriented horizontally in which case they are weakly directional: signals are reasonably well radiated toward or received from all directions with the exception of the direction along the conductor itself; this region is called the antenna blind cone or null.

Both the vertical and dipole antennas are simple in construction and relatively inexpensive. The dipole antenna, which is the basis for most antenna designs, is a balanced component, with equal but opposite voltages and currents applied at its two terminals through a balanced transmission line (or to a coaxial transmission line through a so-called balun). The vertical antenna, on the other hand, is a *monopole* antenna. It is typically connected to the inner conductor of a coaxial transmission line (or a matching network); the shield of the transmission line is connected to ground. In this way, the ground (or any large conductive surface) plays the role of the second conductor of a dipole, thereby forming a complete circuit.<sup>[5]</sup> Since monopole antennas rely on a conductive ground, a so-called grounding structure may be employed to provide a better ground contact to the earth or which itself acts as a ground plane to perform that function regardless of (or in absence of) an actual contact with the earth.



Antennas more complex than the dipole or vertical designs are usually intended to increase the directivity and consequently the gain of the antenna. This can be accomplished in many different ways leading to a plethora of antenna designs. The vast majority of designs are fed with a balanced line (unlike a monopole antenna) and are based on the dipole antenna with additional components (or *elements*) which increase its directionality. Antenna "gain" in this instance describes the concentration of radiated power into a particular solid angle of space, as opposed to the spherically uniform radiation of the ideal radiator. The increased power in the desired direction is at the expense of that in the undesired directions. Power is conserved, and there is no net power increase over that delivered from the power source (the transmitter.)

For instance, a phased array consists of two or more simple antennas which are connected together through an electrical network. This often involves a number of parallel dipole antennas with a certain spacing. Depending on the relative phase introduced by the network, the same combination of dipole antennas can operate as a "broadside array" (directional normal to a line connecting the elements) or as an "end-fire array" (directional along the line connecting the elements). Antenna arrays may employ any basic (omnidirectional or weakly directional) antenna type, such as dipole, loop or slot antennas. These elements are often identical.

However a log-periodic dipole array consists of a number of dipole elements of *different* lengths in order to obtain a somewhat directional antenna having an extremely wide bandwidth: these are frequently used for television reception in fringe areas. The dipole antennas composing it are all considered "active elements" since they are all

electrically connected together (and to the transmission line). On the other hand, a superficially similar dipole array, the Yagi-Uda Antenna (or simply "Yagi"), has only one dipole element with an electrical connection; the other so-called parasitic elements interact with the electromagnetic field in order to realize a fairly directional antenna but one which is limited to a rather narrow bandwidth. The Yagi antenna has similar looking parasitic dipole elements but which act differently due to their somewhat different lengths. There may be a number of so-called "directors" in front of the active element in the direction of propagation, and usually a single (but possibly more) "reflector" on the opposite side of the active element.

Greater directionality can be obtained using beam-forming techniques such as a parabolic reflector or a horn. Since the size of a directional antenna depends on it being large compared to the wavelength, very directional antennas of this sort are mainly feasible at UHF and microwave frequencies. On the other hand, at low frequencies (such as AM broadcast) where a practical antenna must be much smaller than a wavelength, significant directionality isn't even possible. A vertical antenna or loop antenna small compared to the wavelength is typically used, with the main design challenge being that of impedance matching. With a vertical antenna a *loading coil* at the base of the antenna may be employed to cancel the reactive component of impedance; small loop antennas are tuned with parallel capacitors for this purpose.

An antenna lead-in is the transmission line (or *feed line*) which connects the antenna to a transmitter or receiver. The *antenna feed* may refer to all components connecting the antenna to the transmitter or receiver, such as an impedance matching network in addition to the transmission line. In a so-called aperture antenna, such as a horn or parabolic dish, the "feed" may also refer to a basic antenna inside the entire system (normally at the focus of the parabolic dish or at the throat of a horn) which could be considered the one active element in that antenna system. A microwave antenna may also be fed directly from a waveguide in lieu of a (conductive) transmission line.

An antenna counterpoise or ground plane is a structure of conductive material which improves or substitutes for the ground. It may be connected to or insulated from the natural ground. In a monopole antenna, this aids in the function of the natural ground, particularly where variations (or limitations) of the characteristics of the natural ground interfere with its proper function. Such a structure is normally connected to the return connection of an unbalanced transmission line such as the shield of a coaxial cable.

An electromagnetic wave *refractor* in some aperture antennas is a component which due to its shape and position functions to selectively delay or advance portions of the electromagnetic wavefront passing through it. The refractor alters the spatial characteristics of the wave on one side relative to the other side. It can, for instance, bring the wave to a focus or alter the wave front in other ways, generally in order to maximize the directivity of the antenna system. This is the radio equivalent of an optical lens.

An antenna coupling network is a passive network (generally a combination of inductive and capacitive circuit elements) used for impedance matching in between the antenna and the transmitter or receiver. This may be used to improve the standing wave ratio in order to minimize losses in the transmission line and to present the transmitter or receiver with a standard resistive impedance that it expects to see for optimum operation.

## Reciprocity

It is a fundamental property of antennas that the electrical characteristics of an antenna described in the next section, such as gain, radiation pattern, impedance, bandwidth, resonant frequency and polarization, are the same whether the antenna is transmitting or receiving. For example, the "*receiving pattern*" (sensitivity as a function of direction) of an antenna when used for reception is identical to the radiation pattern of the antenna when it is *driven* and functions as a radiator. This is a consequence of the reciprocity theorem of electromagnetics. Therefore in discussions of antenna properties no distinction is usually made between receiving and transmitting terminology, and the antenna can be viewed as either transmitting or receiving, whichever is more convenient.

A necessary condition for the aforementioned reciprocity property is that the materials in the antenna and transmission medium are linear and reciprocal. *Reciprocal* (or *bilateral*) means that the material has the same



response to an electric current or magnetic field in one direction, as it has to the field or current in the opposite direction. Most materials used in antennas meet these conditions, but some microwave antennas use high-tech components such as isolators and circulators, made of nonreciprocal materials such as ferrite or garnet. These can be used to give the antenna a different behavior on receiving than it has on transmitting, which can be useful in applications like radar.

## Parameters

Antennas are characterized by a number of performance measures which a user would be concerned with in selecting or designing an antenna for a particular application. Chief among these relate to the directional characteristics (as depicted in the antenna's *radiation pattern*) and the resulting *gain*. Even in omnidirectional (or weakly directional) antennas, the gain can often be increased by concentrating more of its power in the horizontal directions, sacrificing power radiated toward the sky and ground. The antenna's power gain (or simply "gain") also takes into account the antenna's efficiency, and is often the primary figure of merit.

Resonant antennas are expected to be used around a particular *resonant frequency*; an antenna must therefore be built or ordered to match the frequency range of the intended application. A particular antenna design will present a particular feedpoint impedance. While this may affect the choice of an antenna, an antenna's impedance can also be adapted to the desired impedance level of a system using a matching network while maintaining the other characteristics (except for a possible loss of efficiency).

Although these parameters can be measured in principle, such measurements are difficult and require very specialized equipment. Beyond tuning a transmitting antenna using an SWR meter, the typical user will depend on theoretical predictions based on the antenna design or on claims of a vendor.

An antenna transmits and receives radio waves with a particular polarization which can be reoriented by tilting the axis of the antenna in many (but not all) cases. The physical size of an antenna is often a practical issue, particularly at lower frequencies (longer wavelengths). Highly directional antennas need to be significantly larger than the wavelength. Resonant antennas use a conductor, or a pair of conductors, each of which is about one quarter of the wavelength in length. Antennas that are required to be very small compared to the wavelength sacrifice efficiency and cannot be very directional. Fortunately at higher frequencies (UHF, microwaves) trading off performance to obtain a smaller physical size is usually not required.

## Resonant antennas

While there are broadband designs for antennas, the vast majority of antennas are based on the half-wave dipole which has a particular resonant frequency. At its resonant frequency, the wavelength (figured by dividing the speed of light by the resonant frequency) is slightly over twice the length of the half-wave dipole (thus the name). The quarter-wave vertical antenna consists of one arm of a half-wave dipole, with the other arm replaced by a connection to ground or an equivalent ground plane (or *counterpoise*). A Yagi-Uda array consists of a number of resonant dipole elements, only one of which is directly connected to the transmission line. The quarter-wave elements of a dipole or vertical antenna imitate a series-resonant electrical element, since if they are driven at the resonant frequency a standing wave is created with the peak current at the feed-point and the peak voltage at the far end.

A common misconception is that the ability of a resonant antenna to transmit (or receive) fails at frequencies far from the resonant frequency. The reason a dipole antenna needs to be used at the resonant frequency has to do with the impedance match between the antenna and the transmitter or receiver (and its transmission line). For instance, a dipole using a fairly thin conductor<sup>[6]</sup> will have a purely resistive feedpoint impedance of about 63 ohms at its design frequency. Feeding that antenna with a current of 1 ampere will require 63 volts of RF, and the antenna will radiate 63 watts (ignoring losses) of radio frequency power. If that antenna is driven with 1 ampere at a frequency 20% higher, it will still radiate as efficiently but in order to do that about 200 volts would be required due to the change in the antenna's impedance which is now largely reactive (voltage out of phase with the current). A typical transmitter

would not find that impedance acceptable and would deliver much less than 63 watts to it; the transmission line would be operating at a high (poor) standing wave ratio. But using an appropriate matching network, that large reactive impedance could be converted to a resistive impedance satisfying the transmitter and accepting the available power of the transmitter.

This principle is used to construct vertical antennas substantially shorter than the  $1/4$  wavelength at which the antenna is resonant. By adding an inductance in series with the vertical antenna (a so-called loading coil) the capacitive reactance of this antenna can be cancelled leaving a pure resistance which can then be matched to the transmission line. Sometimes the resulting resonant frequency of such a system (antenna plus matching network) is described using the construct of "electrical length" and the use of a shorter antenna at a lower frequency than its resonant frequency is termed "electrical lengthening". For example, at 30 MHz (wavelength = 10 meters) a true resonant monopole would be almost 2.5 meters ( $1/4$  wavelength) long, and using an antenna only 1.5 meters tall would require the addition of a loading coil. Then it may be said that the coil has "lengthened" the antenna to achieve an "electrical length" of 2.5 meters, that is,  $1/4$  wavelength at 30 MHz where the combined system now resonates. However, the resulting resistive impedance achieved will be quite a bit lower than the impedance of a resonant monopole, likely requiring further impedance matching. In addition to a lower radiation resistance, the reactance becomes higher as the antenna size is reduced, and the resonant circuit formed by the antenna and the tuning coil has a Q factor that rises and eventually causes the bandwidth of the antenna to be inadequate for the signal being transmitted. This is the major factor that sets the size of antennas at 1 MHz and lower frequencies.

### Current and voltage distribution

The antenna conductors have the lowest feed-point impedance at the resonant frequency where they are just under  $1/4$  wavelength long; two such conductors in line fed differentially thus realizes the familiar "half-wave dipole". When fed with an RF current at the resonant frequency, the quarter wave element contains a standing wave with the voltage and current largely (but not exactly) in phase quadrature, as would be obtained using a quarter wave stub of transmission line. The current reaches a minimum at the end of the element (where it has nowhere to go!) and is maximum at the feed-point. The voltage, on the other hand, is the greatest at the end of the conductor and reaches a minimum (but not zero) at the feedpoint. Making the conductor shorter or longer than  $1/4$  wavelength means that the voltage pattern reaches its minimum somewhere beyond the feed-point, so that the feed-point has a higher voltage and thus sees a higher impedance, as we have noted. Since that voltage pattern is almost in phase quadrature with the current, the impedance seen at the feed-point is not only much higher but mainly reactive.

It can be seen that if such an element is resonant at  $f_0$  to produce such a standing wave pattern, then feeding that element with  $3f_0$  (whose wavelength is  $1/3$  that of  $f_0$ ) will lead to a standing wave pattern in which the voltage is likewise a minimum at the feed-point (and the current at a maximum there). Thus, an antenna element is *also* resonant when its length is  $3/4$  of a wavelength ( $3/2$  wavelength for a complete dipole). This is true for all odd multiples of  $1/4$  wavelength, where the feed-point impedance is purely resistive, though larger than the resistive impedance of the  $1/4$  wave element. Although such an antenna is resonant and works perfectly well at the higher frequency, the antenna radiation pattern is also altered compared to the half-wave dipole.

The use of a monopole or dipole at odd multiples of the fundamental resonant frequency, however, does *not* extend to even multiples (thus a  $1/2$  wavelength monopole or 1 wavelength dipole). Now the voltage standing wave is at its *peak* at the feed-point, while that of the current (which must be zero at the end of the conductor) is at a minimum (but not exactly zero). The antenna is *anti-resonant* at this frequency. Although the reactance at the feedpoint can be cancelled using such an element length, the feed-point impedance is very high, and is highly dependent on the diameter of the conductor (which makes only a small difference at the actual resonant frequency). Such an antenna does not match the much lower characteristic impedance of available transmission lines, and is generally not used. However some equipment where transmission lines are not involved which desire a high driving point impedance may take advantage of this anti-resonance.



## Bandwidth

Although a resonant antenna has a purely resistive feed-point impedance at a particular frequency, many (if not most) applications require using an antenna over a range of frequencies. An antenna's *bandwidth* specifies the range of frequencies over which its performance does not suffer due to a poor impedance match. Also in the case of a Yagi-Uda array, the use of the antenna very far away from its design frequency reduces the antenna's directivity, thus reducing the usable bandwidth regardless of impedance matching.

Except for the latter concern, the resonant frequency of a resonant antenna can always be altered by adjusting a suitable matching network. To do this efficiently one would require remotely adjusting a matching network at the site of the antenna, since simply adjusting a matching network at the transmitter (or receiver) would leave the transmission line with a poor standing wave ratio.

Instead, it is often desired to have an antenna whose impedance does not vary so greatly over a certain bandwidth. It turns out that the amount of reactance seen at the terminals of a resonant antenna when the frequency is shifted, say, by 5%, depends very much on the diameter of the conductor used. A long thin wire used as a half-wave dipole (or quarter wave monopole) will have a reactance significantly greater than the resistive impedance it has at resonance, leading to a poor match and generally unacceptable performance. Making the element using a tube of a diameter perhaps 1/50 of its length, however, results in a reactance at this altered frequency which is not so great, and a much less serious mismatch which will only modestly damage the antenna's net performance. Thus rather thick tubes are typically used for the solid elements of such antennas, including Yagi-Uda arrays.

Rather than just using a thick tube, there are similar techniques used to the same effect such as replacing thin wire elements with *cages* to simulate a thicker element. This widens the bandwidth of the resonance. On the other hand, amateur radio antennas need to operate over several bands which are widely separated from each other. This can often be accomplished simply by connecting resonant elements for the different bands in parallel. Most of the transmitter's power will flow into the resonant element while the others present a high (reactive) impedance and draw little current from the same voltage. A popular solution uses so-called *traps* consisting of parallel resonant circuits which are strategically placed in breaks along each antenna element. When used at one particular frequency band the trap presents a very high impedance (parallel resonance) effectively truncating the element at that length, making it a proper resonant antenna. At a lower frequency the trap allows the full length of the element to be employed, albeit with a shifted resonant frequency due to the inclusion of the trap's net reactance at that lower frequency.

The bandwidth characteristics of a resonant antenna element can be characterized according to its  $Q$ , just as one uses to characterize the sharpness of an L-C resonant circuit. However it is often assumed that there is an advantage in an antenna having a high  $Q$ . After all,  $Q$  is short for "quality factor" and a low  $Q$  typically signifies excessive loss (due to unwanted resistance) in a resonant L-C circuit. However this understanding does not apply to resonant antennas where the resistance involved is the radiation resistance, a desired quantity which removes energy from the resonant element in order to radiate it (the purpose of an antenna, after all!). The  $Q$  is a measure of the ratio of reactance to resistance, so with a fixed radiation resistance (an element's radiation resistance is almost independent of its diameter) a greater reactance off-resonance corresponds to the poorer bandwidth of a very thin conductor. The  $Q$  of such a narrowband antenna can be as high as 15. On the other hand a thick element presents less reactance at an off-resonant frequency, and consequently a  $Q$  as low as 5. These two antennas will perform equivalently at the resonant frequency, but the second antenna will perform over a bandwidth 3 times as wide as the "hi- $Q$ " antenna consisting of a thin conductor.

## Gain

Gain is a parameter which measures the degree of directivity of the antenna's radiation pattern. A high-gain antenna will preferentially radiate in a particular direction. Specifically, the *antenna gain*, or *power gain* of an antenna is defined as the ratio of the intensity (power per unit surface) radiated by the antenna in the direction of its maximum output, at an arbitrary distance, divided by the intensity radiated at the same distance by a hypothetical isotropic antenna.

The gain of an antenna is a passive phenomenon - power is not added by the antenna, but simply redistributed to provide more radiated power in a certain direction than would be transmitted by an isotropic antenna. An antenna designer must take into account the application for the antenna when determining the gain. High-gain antennas have the advantage of longer range and better signal quality, but must be aimed carefully in a particular direction. Low-gain antennas have shorter range, but the orientation of the antenna is relatively inconsequential. For example, a dish antenna on a spacecraft is a high-gain device that must be pointed at the planet to be effective, whereas a typical Wi-Fi antenna in a laptop computer is low-gain, and as long as the base station is within range, the antenna can be in any orientation in space. It makes sense to improve horizontal range at the expense of reception above or below the antenna.<sup>[7]</sup>

In practice, the half-wave dipole is taken as a reference instead of the isotropic radiator. The gain is then given in **dBd** (decibels over dipole):

NOTE: **0 dBd = 2.15 dBi**. It is vital in expressing gain values that the reference point be included. Failure to do so can lead to confusion and error.

## Effective area or aperture

The *effective area* or effective aperture of a receiving antenna expresses the portion of the power of a passing electromagnetic wave which it delivers to its terminals, expressed in terms of an equivalent area. For instance, if a radio wave passing a given location has a flux of 1 pW / m<sup>2</sup> (10<sup>-12</sup> watts per square meter) and an antenna has an effective area of 12 m<sup>2</sup>, then the antenna would deliver 12 pW of RF power to the receiver (30 microvolts rms at 75 ohms). Since the receiving antenna is not equally sensitive to signals received from all directions, the effective area is a function of the direction to the source.

Due to reciprocity (discussed above) the gain of an antenna used for transmitting must be proportional to its effective area when used for receiving. Consider an antenna with no loss, that is, one whose electrical efficiency is 100%. It can be shown that its effective area averaged over all directions must be equal to  $\lambda^2/4\pi$ , the wavelength squared divided by  $4\pi$ . Gain is defined such that the average gain over all directions for an antenna with 100% electrical efficiency is equal to 1. Therefore the effective area  $A_{\text{eff}}$  in terms of the gain  $G$  in a given direction is given by:

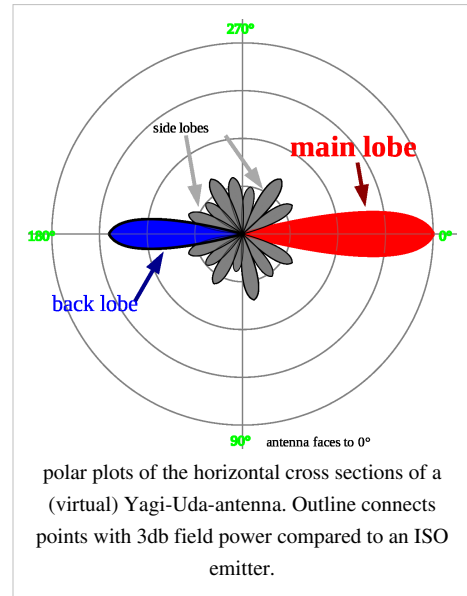
$$A_{\text{eff}} = \frac{\lambda^2}{4\pi} G$$

For an antenna with an efficiency of less than 100%, both the effective area and gain are reduced by that same amount. Therefore the above relationship between gain and effective area still holds. These are thus two different ways of expressing the same quantity.  $A_{\text{eff}}$  is especially convenient when computing the power that would be received by an antenna of a specified gain, as illustrated by the above example.

## Radiation pattern

The radiation pattern of an antenna is a plot of the relative field strength of the radio waves emitted by the antenna at different angles. It is typically represented by a three dimensional graph, or polar plots of the horizontal and vertical cross sections. The pattern of an ideal isotropic antenna, which radiates equally in all directions, would look like a sphere. Many nondirectional antennas, such as monopoles and dipoles, emit equal power in all horizontal directions, with the power dropping off at higher and lower angles; this is called an omnidirectional pattern and when plotted looks like a torus or donut.

The radiation of many antennas shows a pattern of maxima or "*lobes*" at various angles, separated by "*nulls*", angles where the radiation falls to zero. This is because the radio waves emitted by different parts of the antenna typically interfere, causing maxima at angles where the radio waves arrive at distant points in phase, and zero radiation at other angles where the radio waves arrive out of phase. In a directional antenna designed to project radio waves in a particular direction, the lobe in that direction is designed larger than the others and is called the "*main lobe*". The other lobes usually represent unwanted radiation and are called "*sidelobes*". The axis through the main lobe is called the "*principal axis*" or "*boresight axis*".



## Field regions

The space surrounding an antenna can be divided into three concentric regions: the reactive near-field, the radiating near-field (Fresnell region) and the far-field (Fraunhofer) regions. These regions are useful to identify the field structure in each, although there are no precise boundaries.

In the far-field region, we are far enough from the antenna to neglect its size and shape. We can assume that the electromagnetic wave is purely a radiating plane wave (electric and magnetic fields are in phase and perpendicular to each other and to the direction of propagation). This simplifies the mathematical analysis of the radiated field.

## Impedance

As an electro-magnetic wave travels through the different parts of the antenna system (radio, feed line, antenna, free space) it may encounter differences in impedance ( $E/H$ ,  $V/I$ , etc.). At each interface, depending on the impedance match, some fraction of the wave's energy will reflect back to the source,<sup>[8]</sup> forming a standing wave in the feed line. The ratio of maximum power to minimum power in the wave can be measured and is called the standing wave ratio (**SWR**). A SWR of 1:1 is ideal. A SWR of 1.5:1 is considered to be marginally acceptable in low power applications where power loss is more critical, although an SWR as high as 6:1 may still be usable with the right equipment. Minimizing impedance differences at each interface (impedance matching) will reduce SWR and maximize power transfer through each part of the antenna system.

Complex impedance of an antenna is related to the electrical length of the antenna at the wavelength in use. The impedance of an antenna can be matched to the feed line and radio by adjusting the impedance of the feed line, using the feed line as an impedance transformer. More commonly, the impedance is adjusted at the load (see below) with an antenna tuner, a balun, a matching transformer, matching networks composed of inductors and capacitors, or matching sections such as the gamma match.

## Efficiency

*Efficiency* of a transmitting antenna is the ratio of power actually radiated (in all directions) to the power absorbed by the antenna terminals. The power supplied to the antenna terminals which is not radiated is converted into heat. This is usually through loss resistance in the antenna's conductors, but can also be due to dielectric or magnetic core losses in antennas (or antenna systems) using such components. Such loss effectively robs power from the transmitter, requiring a stronger transmitter in order to transmit a signal of a given strength.

For instance, if a transmitter delivers 100 W into an antenna having an efficiency of 80%, then the antenna will radiate 80 W as radio waves and produce 20 W of heat. In order to radiate 100 W of power, one would need to use a transmitter capable of supplying 125 W to the antenna. Note that antenna efficiency is a separate issue from impedance matching, which may also reduce the amount of power radiated using a given transmitter. If an SWR meter reads 150 W of incident power and 50 W of reflected power, that means that 100 W have actually been absorbed by the antenna (ignoring transmission line losses). How much of that power has actually been radiated cannot be directly determined through electrical measurements at (or before) the antenna terminals, but would require (for instance) careful measurement of field strength. Fortunately the loss resistance of antenna conductors such as aluminum rods can be calculated and the efficiency of an antenna using such materials predicted.

However loss resistance will generally affect the feedpoint impedance, adding to its resistive (real) component. That resistance will consist of the sum of the radiation resistance  $R_r$  and the loss resistance  $R_{\text{loss}}$ . If an rms current  $I$  is delivered to the terminals of an antenna, then a power of  $I^2 R_r$  will be radiated and a power of  $I^2 R_{\text{loss}}$  will be lost as heat. Therefore the efficiency of an antenna is equal to  $R_r / (R_r + R_{\text{loss}})$ . Of course only the total resistance  $R_r + R_{\text{loss}}$  can be directly measured.

According to reciprocity, the efficiency of an antenna used as a receiving antenna is identical to the efficiency as defined above. The power that an antenna will deliver to a receiver (with a proper impedance match) is reduced by the same amount. In some receiving applications, the very inefficient antennas may have little impact on performance. At low frequencies, for example, atmospheric or man-made noise can mask antenna inefficiency. For example, CCIR Rep. 258-3 indicates man-made noise in a residential setting at 40 MHz is about 28 dB above the thermal noise floor. Consequently, an antenna with a 20 dB loss (due to inefficiency) would have little impact on system noise performance. The loss within the antenna will affect the intended signal and the noise/interference identically, leading to no reduction in signal to noise ratio (SNR).

This is fortunate, since antennas at lower frequencies which are not rather large (a good fraction of a wavelength in size) are inevitably inefficient (due to the small radiation resistance  $R_r$  of small antennas). Most AM broadcast radios (except for car radios) take advantage of this principle by including a small loop antenna for reception which has an extremely poor efficiency. Using such an inefficient antenna at this low frequency (530–1650 kHz) thus has little effect on the receiver's net performance, but simply requires greater amplification by the receiver's electronics. Contrast this tiny component to the massive and very tall towers used at AM broadcast stations for transmitting at the very same frequency, where every percentage point of reduced antenna efficiency entails a substantial cost.

The definition of antenna gain or *power gain* already includes the effect of the antenna's efficiency. Therefore if one is trying to radiate a signal toward a receiver using a transmitter of a given power, one need only compare the gain of various antennas rather than considering the efficiency as well. This is likewise true for a receiving antenna at very high (especially microwave) frequencies, where the point is to receive a signal which is strong compared to the receiver's noise temperature. However in the case of a directional antenna used for receiving signals with the intention of *rejecting* interference from different directions, one is no longer concerned with the antenna efficiency, as discussed above. In this case, rather than quoting the antenna gain, one would be more concerned with the *directive gain* which does *not* include the effect of antenna (in)efficiency. The directive gain of an antenna can be computed from the published gain divided by the antenna's efficiency.

## Polarization

The *polarization* of an antenna is the orientation of the electric field (E-plane) of the radio wave with respect to the Earth's surface and is determined by the physical structure of the antenna and by its orientation. It has nothing in common with antenna directionality terms: "horizontal", "vertical", and "circular". Thus, a simple straight wire antenna will have one polarization when mounted vertically, and a different polarization when mounted horizontally. "Electromagnetic wave polarization filters" are structures which can be employed to act directly on the electromagnetic wave to filter out wave energy of an undesired polarization and to pass wave energy of a desired polarization.

Reflections generally affect polarization. For radio waves the most important reflector is the ionosphere - signals which reflect from it will have their polarization changed unpredictably. For signals which are reflected by the ionosphere, polarization cannot be relied upon. For line-of-sight communications for which polarization can be relied upon, it can make a large difference in signal quality to have the transmitter and receiver using the same polarization; many tens of dB difference are commonly seen and this is more than enough to make the difference between reasonable communication and a broken link.

Polarization is largely predictable from antenna construction but, especially in directional antennas, the polarization of side lobes can be quite different from that of the main propagation lobe. For radio antennas, polarization corresponds to the orientation of the radiating element in an antenna. A vertical omnidirectional WiFi antenna will have vertical polarization (the most common type). An exception is a class of elongated waveguide antennas in which vertically placed antennas are horizontally polarized. Many commercial antennas are marked as to the polarization of their emitted signals.

Polarization is the sum of the E-plane orientations over time projected onto an imaginary plane perpendicular to the direction of motion of the radio wave. In the most general case, polarization is elliptical, meaning that the polarization of the radio waves varies over time. Two special cases are linear polarization (the ellipse collapses into a line) and circular polarization (in which the two axes of the ellipse are equal). In linear polarization the antenna compels the electric field of the emitted radio wave to a particular orientation. Depending on the orientation of the antenna mounting, the usual linear cases are horizontal and vertical polarization. In circular polarization, the antenna continuously varies the electric field of the radio wave through all possible values of its orientation with regard to the Earth's surface. Circular polarizations, like elliptical ones, are classified as right-hand polarized or left-hand polarized using a "thumb in the direction of the propagation" rule. Optical researchers use the same rule of thumb, but pointing it in the direction of the emitter, not in the direction of propagation, and so are opposite to radio engineers' use.

In practice, regardless of confusing terminology, it is important that linearly polarized antennas be matched, lest the received signal strength be greatly reduced. So horizontal should be used with horizontal and vertical with vertical. Intermediate matchings will lose some signal strength, but not as much as a complete mismatch. Transmitters mounted on vehicles with large motional freedom commonly use circularly polarized antennas so that there will never be a complete mismatch with signals from other sources.

## Impedance matching

Maximum power transfer requires matching the impedance of an antenna system (as seen looking into the transmission line) to the complex conjugate of the impedance of the receiver or transmitter. In the case of a transmitter, however, the desired matching impedance might not correspond to the dynamic output impedance of the transmitter as analyzed as a source impedance but rather the design value (typically 50 ohms) required for efficient and safe operation of the transmitting circuitry. The intended impedance is normally resistive but a transmitter (and some receivers) may have additional adjustments to cancel a certain amount of reactance in order to "tweak" the match. When a transmission line is used in between the antenna and the transmitter (or receiver) one generally would like an antenna system whose impedance is resistive and near the characteristic impedance of that transmission line

in order to minimize the standing wave ratio (SWR) and the increase in transmission line losses it entails, in addition to supplying a good match at the transmitter or receiver itself.

Antenna tuning generally refers to cancellation of any reactance seen at the antenna terminals, leaving only a resistive impedance which might or might not be exactly the desired impedance (that of the transmission line). Although an antenna may be designed to have a purely resistive feedpoint impedance (such as a dipole 97% of a half wavelength long) this might not be exactly true at the frequency that it is eventually used at. In some cases the physical length of the antenna can be "trimmed" to obtain a pure resistance. On the other hand, the addition of a series inductance or parallel capacitance can be used to cancel a residual capacitive or inductive reactance, respectively.

In some cases this is done in a more extreme manner, not simply to cancel a small amount of residual reactance, but to resonate an antenna whose resonance frequency is quite different than the intended frequency of operation. For instance, a "whip antenna" can be made significantly shorter than 1/4 wavelength long, for practical reasons, and then resonated using a so-called loading coil. This physically large inductor at the base of the antenna has an inductive reactance which is the opposite of the capacitive reactance that such a vertical antenna has at the desired operating frequency. The result is a pure resistance seen at feedpoint of the loading coil; unfortunately that resistance is somewhat lower than would be desired to match commercial coax.

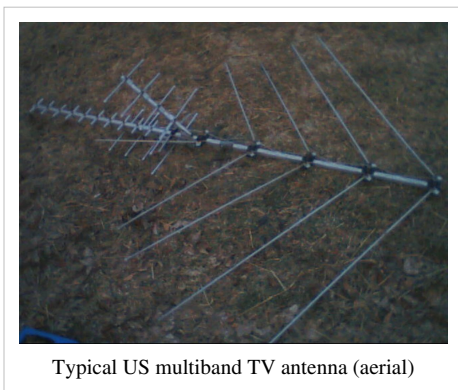
So an additional problem beyond canceling the unwanted reactance is of matching the remaining resistive impedance to the characteristic impedance of the transmission line. In principle this can always be done with a transformer, however the turns ratio of a transformer is not adjustable. A general matching network with at least two adjustments can be made to correct both components of impedance. Matching networks using discrete inductors and capacitors will have losses associated with those components, and will have power restrictions when used for transmitting. Avoiding these difficulties, commercial antennas are generally designed with fixed matching elements or feeding strategies to get an approximate match to standard coax, such as 50 or 75 Ohms. Antennas based on the dipole (rather than vertical antennas) should include a balun in between the transmission line and antenna element, which may be integrated into any such matching network.

Another extreme case of impedance matching occurs when using a small loop antenna (usually, but not always, for receiving) at a relatively low frequency where it appears almost as a pure inductor. Resonating such an inductor with a capacitor at the frequency of operation not only cancels the reactance but greatly magnifies the very small radiation resistance of such a loop. This is implemented in most AM broadcast receivers, with a small ferrite loop antenna resonated by a capacitor which is varied along with the receiver tuning in order to maintain resonance over the AM broadcast band

## Basic antenna models

There are many variations of antennas. Below are a few basic models. More can be found in Category:Radio frequency antenna types.

- The isotropic radiator is a purely theoretical antenna that radiates equally in all directions. It is considered to be a point in space with no dimensions and no mass. This antenna cannot physically exist, but is useful as a theoretical model for comparison with all other antennas. Most antennas' gains are measured with reference to an isotropic radiator, and are rated in dBi (decibels with respect to an isotropic radiator).



Typical US multiband TV antenna (aerial)

- The dipole antenna is simply two wires pointed in opposite directions arranged either horizontally or vertically, with one end of each wire connected to the radio and the other end hanging free in space. Since this is the

simplest practical antenna, it is also used as a reference model for other antennas; gain with respect to a dipole is labeled as dBd. Generally, the dipole is considered to be omnidirectional in the plane perpendicular to the axis of the antenna, but it has deep nulls in the directions of the axis. Variations of the dipole include the folded dipole, the half wave antenna, the ground plane antenna, the whip, and the J-pole.

- The Yagi-Uda antenna is a directional variation of the dipole with parasitic elements added which are functionality similar to adding a reflector and lenses (directors) to focus a filament light bulb.
- The random wire antenna is simply a very long (at least one quarter wavelength) wire with one end connected to the radio and the other in free space, arranged in any way most convenient for the space available. Folding will reduce effectiveness and make theoretical analysis extremely difficult. (The added length helps more than the folding typically hurts.) Typically, a random wire antenna will also require an antenna tuner, as it might have a random impedance that varies non-linearly with frequency.
- The horn antenna is used where high gain is needed, the wavelength is short (microwave) and space is not an issue. Horns can be narrow band or wide band, depending on their shape. A horn can be built for any frequency, but horns for lower frequencies are typically impractical. Horns are also frequently used as reference antennas.
- The parabolic antenna consists of an active element at the focus of a parabolic reflector to reflect the waves into a plane wave. Like the horn it is used for high gain, microwave applications, such as satellite dishes.
- The patch antenna consists mainly of a square conductor mounted over a groundplane. Another example of a planar antenna is the tapered slot antenna (TSA), as the Vivaldi-antenna.

## Practical antennas

Although any circuit can radiate if driven with a signal of high enough frequency, most practical antennas are specially designed to radiate efficiently at a particular frequency. An example of an inefficient antenna is the simple Hertzian dipole antenna, which radiates over wide range of frequencies and is useful for its small size. A more efficient variation of this is the half-wave dipole, which radiates with high efficiency when the signal wavelength is twice the electrical length of the antenna.

One of the goals of antenna design is to minimize the reactance of the device so that it appears as a resistive load. An "antenna inherent reactance" includes not only the distributed reactance of the active antenna but also the natural reactance due to its location and surroundings (as for example, the capacity relation inherent in the position of the active antenna relative to ground). Reactance can be eliminated by operating the antenna at its resonant frequency, when its capacitive and inductive reactances are equal and opposite, resulting in a net zero reactive current. If this is not possible, compensating inductors or capacitors can instead be added to the antenna to cancel its reactance as far as the source is concerned.

Once the reactance has been eliminated, what remains is a pure resistance, which is the sum of two parts: the ohmic resistance of the conductors, and the radiation resistance. Power absorbed by the ohmic resistance becomes waste heat, and that absorbed by the radiation resistance becomes radiated electromagnetic energy. The greater the ratio of radiation resistance to ohmic resistance, the more efficient the antenna.





### Effect of ground

Antennas are typically used in an environment where other objects are present that may have an effect on their performance. Height above ground has a very significant effect on the radiation pattern of some antenna types.

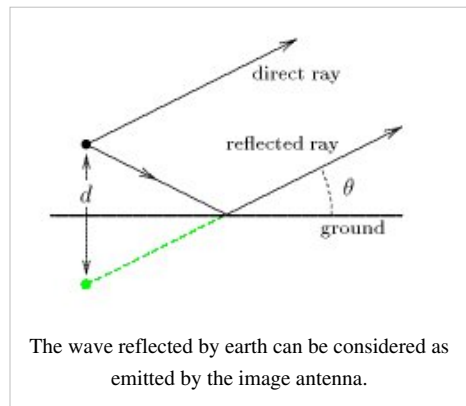
At frequencies used in antennas, the ground behaves mainly as a dielectric. The conductivity of ground at these frequencies is negligible. When an electromagnetic wave arrives at the surface of an object, two waves are created: one enters the dielectric and the other is reflected. If the object is a conductor, the transmitted wave is negligible and the reflected wave has almost the same amplitude as the incident one. When the object is a dielectric, the fraction reflected depends (among other things) on the angle of incidence. When the angle of incidence is small (that is, the wave arrives almost perpendicularly) most of the energy traverses the surface and very little is reflected. When the angle of incidence is near 90° (grazing incidence) almost all the wave is reflected.

Most of the electromagnetic waves emitted by an antenna to the ground below the antenna at moderate (say < 60°) angles of incidence enter the earth and are absorbed (lost). But waves emitted to the ground at grazing angles, far from the antenna, are almost totally reflected. At grazing angles, the ground behaves as a mirror. Quality of reflection depends on the nature of the surface. When the irregularities of the surface are smaller than the wavelength, reflection is good.

This means that the receptor "sees" the real antenna and, under the ground, the image of the antenna reflected by the ground. If the ground has irregularities, the image will appear fuzzy.

If the receiver is placed at some height above the ground, waves reflected by ground will travel a little longer distance to arrive to the receiver than direct waves. The distance will be the same only if the receiver is close to ground.

In the drawing at right, the angle has been drawn  $\theta$  far bigger than in reality. The distance between the antenna and its image is  $d$ .



The situation is a bit more complex because the reflection of electromagnetic waves depends on the polarization of the incident wave. As the refractive index of the ground (average value  $\approx 2$ ) is bigger than the refractive index of the air ( $\approx 1$ ), the direction of the component of the electric field parallel to the ground inverts at the reflection. This is equivalent to a phase shift of  $\pi$  radians or 180°. The vertical component of the electric field reflects without changing direction. This sign inversion of the parallel component and the non-inversion of the perpendicular component would also happen if the ground were a good electrical conductor.

This means that a receiving antenna "sees" the image antenna with the current in the same direction if the antenna is vertical or with the current inverted if the antenna is horizontal.

For a vertical polarized emission antenna the far electric field of the electromagnetic wave produced by the direct ray plus the reflected ray is:

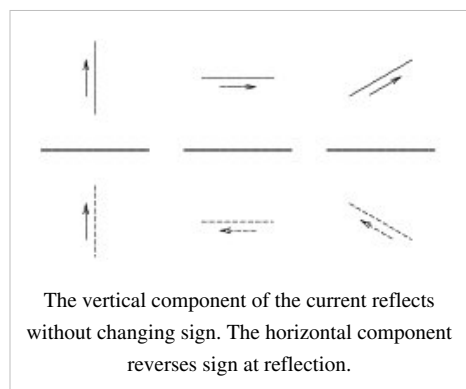
$$|E_{\perp}| = 2 |E_{\theta_1}| \left| \cos \left( \frac{kd}{2} \sin \theta \right) \right|$$

The sign inversion for the parallel field case just changes a cosine to a sine:

$$|E_{\parallel}| = 2 |E_{\theta_1}| \left| \sin \left( \frac{kd}{2} \sin \theta \right) \right|$$

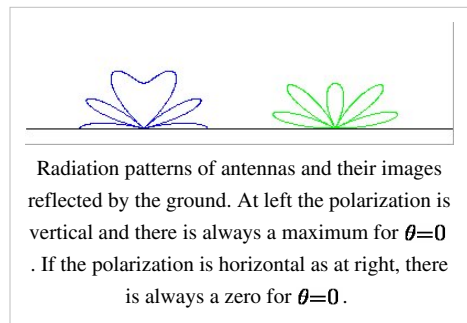
In these two equations:

- $E_{\theta_1}$  is the electrical field radiated by the antenna if there were no ground.



- $k = \frac{2\pi}{\lambda}$  is the wave number.
- $\lambda$  is the wave length.
- $d$  is the distance between antenna and its image (twice the height of the center of the antenna).

For emitting and receiving antennas situated near the ground (in a building or on a mast) far from each other, distances traveled by direct and reflected rays are nearly the same. There is no induced phase shift. If the emission is polarized vertically, the two fields (direct and reflected) add and there is maximum of received signal. If the emission is polarized horizontally, the two signals subtract and the received signal is minimum. This is depicted in the image at right. In the case of vertical polarization, there is always a maximum at earth level (left pattern). For horizontal polarization, there is always a minimum at earth level. Note that in these drawings the ground is considered as a perfect mirror, even for low angles of incidence. In these drawings, the distance between the antenna and its image is just a few wavelengths. For greater distances, the number of lobes increases.

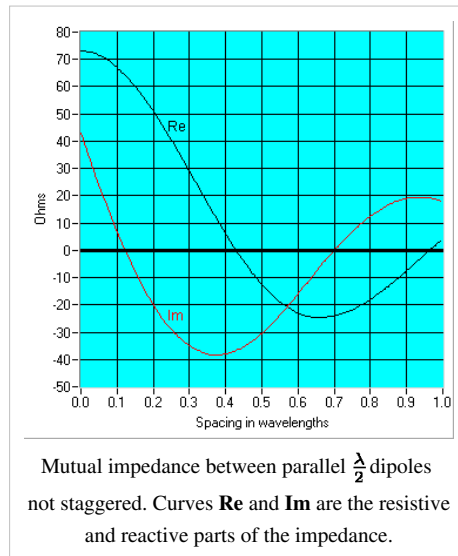


Note that the situation is different—and more complex—if reflections in the ionosphere occur. This happens over very long distances (thousands of kilometers). There is not a direct ray but several reflected rays that add with different phase shifts.

This is the reason why almost all public address radio emissions have vertical polarization. As public users are near ground, horizontal polarized emissions would be poorly received. Observe household and automobile radio receivers. They all have vertical antennas or horizontal ferrite antennas for vertical polarized emissions. In cases where the receiving antenna must work in any position, as in mobile phones, the emitter and receivers in base stations use circular polarized electromagnetic waves.

Classical (analog) television emissions are an exception. They are almost always horizontally polarized, because the presence of buildings makes it unlikely that a good emitter antenna image will appear. However, these same buildings reflect the electromagnetic waves and can create ghost images. Using horizontal polarization, reflections are attenuated because of the low reflection of electromagnetic waves whose magnetic field is parallel to the dielectric surface near the Brewster's angle. Vertically polarized analog television has been used in some rural areas. In digital terrestrial television reflections are less obtrusive, due to the inherent robustness of digital signalling and built-in error correction.

## Mutual impedance and interaction between antennas



Current circulating in any antenna induces currents in all others. One can postulate a **mutual impedance**  $Z_{12}$  between two antennas that has the same significance as the  $j\omega M$  in ordinary coupled inductors. The mutual impedance  $Z_{12}$  between two antennas is defined as:

$$Z_{12} = \frac{v_2}{i_1}$$

where  $i_1$  is the current flowing in antenna 1 and  $v_2$  is the voltage that would have to be applied to antenna 2—with antenna 1 removed—to produce the current in the antenna 2 that was produced by antenna 1.

From this definition, the currents and voltages applied in a set of coupled antennas are:

$$\begin{aligned} v_1 &= i_1 Z_{11} + i_2 Z_{12} + \dots + i_n Z_{1n} \\ v_2 &= i_1 Z_{21} + i_2 Z_{22} + \dots + i_n Z_{2n} \\ &\vdots \\ v_n &= i_1 Z_{n1} + i_2 Z_{n2} + \dots + i_n Z_{nn} \end{aligned}$$

where:

- $v_i$  is the voltage applied to the antenna  $i$
- $Z_{ii}$  is the impedance of antenna  $i$
- $Z_{ij}$  is the mutual impedance between antennas  $i$  and  $j$

Note that, as is the case for mutual inductances,

$$Z_{ij} = Z_{ji}$$

This is a consequence of Lorentz reciprocity. If some of the elements are not fed (there is a short circuit instead a feeder cable), as is the case in television antennas (Yagi-Uda antennas), the corresponding  $v_i$  are zero. Those elements are called parasitic elements. Parasitic elements are unpowered elements that either reflect or absorb and reradiate RF energy.

In some geometrical settings, the mutual impedance between antennas can be zero. This is the case for crossed dipoles used in circular polarization antennas.

## Antenna gallery

### Antennas and antenna arrays



A Yagi-Uda beam antenna.



A multi-band rotary directional antenna for amateur radio use.



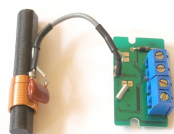
Rooftop TV antenna. It is actually three Yagi antennas. The longest elements are for the low band, while the medium and short elements are for the high and UHF band.



A terrestrial microwave radio antenna array.



Wire dipole antenna using open-wire ladder line feedline for amateur radio use.



Low cost LF time signal receiver, antenna (left) and receiver



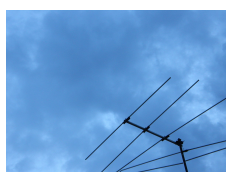
Examples of US 136-174 MHz base station antennas.



Rotatable log-periodic array for VHF and UHF.



Shortwave antennas in Delano, California.



An old VHF-band Yagi-type television antenna.



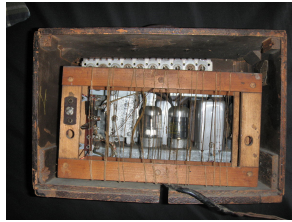
A T2FD broadband antenna, covering the 5-30 MHz band.



A US multiband "aerial" TV antenna.



"Rabbit ears" antenna



AM loop antenna

### Antennas and supporting structures



A building rooftop supporting numerous dish and sectored mobile telecommunications antennas (Doncaster, Victoria, Australia).



A water tower in Palmerston, Northern Territory with radio broadcasting and communications antennas.

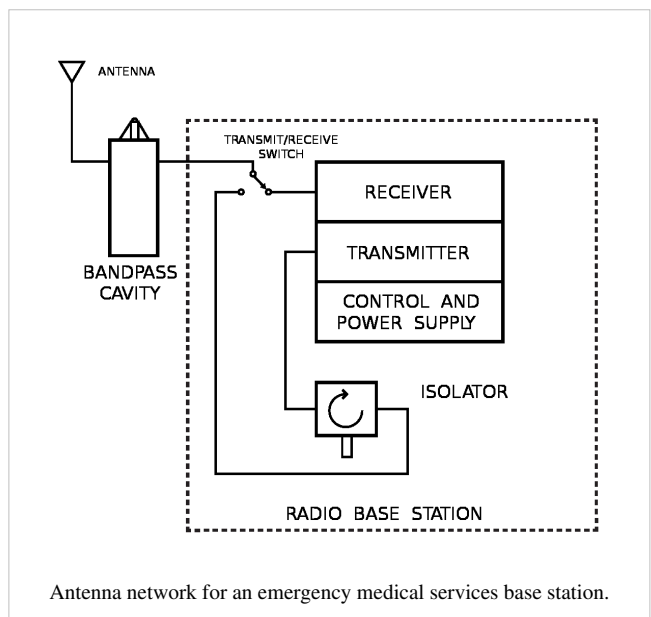
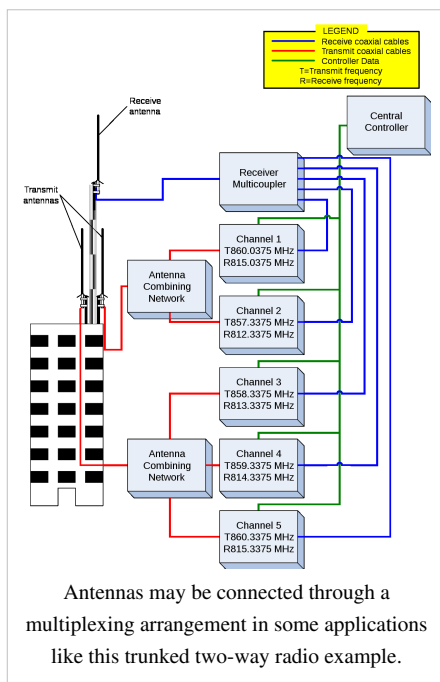


A three-sector telephone site in Mexico City.



Telephone site concealed as a palm tree.

### Diagrams as part of a system



## Notes

- [1] In the context of electrical engineering and physics, the plural of *antenna* is *antennas*, and it has been this way since about 1950 (or earlier), when a cornerstone textbook in this field, *Antennas*, was published by the physicist and electrical engineer John D. Kraus of The Ohio State University. Besides in the title, Dr. Kraus noted this in a footnote on the first page of his book. Insects may have "antennae", but this form is not used in the context of electronics or physics.
- [2] <http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?reload=true&punumber=8>
- [3] "*Salvan: Cradle of Wireless, How Marconi Conducted Early Wireless Experiments in the Swiss Alps*", Fred Gardiol & Yves Fournier, *Microwave Journal*, February 2006, pp. 124-136.
- [4] "Media Advisory: Apply Now to Attend the ALMA Observatory Inauguration" (<http://www.eso.org/public/announcements/ann12092/>). *ESO Announcement*. Retrieved 4 December 2012.
- [5] Tesla said during the development of radio that "*One of the terminals of the source would be connected to Earth [as a electric ground connection ...] the other to an insulated body of large surface*". For more information, see "*On Light and Other High Frequency Phenomena*" (<http://www.tfcbooks.com/tesla/1893-02-24.htm>). Delivered before the Franklin Institute, Philadelphia, February 1893, and before the National Electric Light Association, St. Louis, Missouri, March 1893.
- [6] This example assumes a length to diameter ratio of 1000.
- [7] "Guide to Wi-Fi Wireless Network Antenna Selection." (<http://networkbits.net/wireless-printing/wireless-network-antenna-guide/>). NetworkBits.net. Archived (<http://web.archive.org/web/20080305182010/http://networkbits.net/wireless-printing/wireless-network-antenna-guide/>) from the original on 5 March 2008. Retrieved April 8, 2008.
- [8] Impedance is caused by the same physics as refractive index in optics, although impedance effects are typically one dimensional, where effects of refractive index is three dimensional.

## References

### General references

- Antenna Theory (3rd edition), by C. Balanis, Wiley, 2005, ISBN 0-471-66782-X;
- Antenna Theory and Design (2nd edition), by W. Stutzman and G. Thiele, Wiley, 1997, ISBN 0-471-02590-9;
- Antennas (3rd edition), by J. Kraus and R. Marhefka, McGraw-Hill, 2001, ISBN 0-07-232103-2;
- Antennenbuch, by Karl Rothammel, publ. Franck'sche Verlagshandlung Stuttgart, 1991, ISBN 3-440-05853-0; other editions (<http://www.worldcat.org/oclc/65969707?tab=editions>) (in German)
- Antennas for portable Devices (<http://www1.i2r.a-star.edu.sg/~chenzn>), Zhi Ning Chen (edited), John Wiley & Sons in March 2007
- Broadband Planar Antennas: Design and Applications, Zhi Ning Chen and M. Y. W. Chia, John Wiley & Sons in February 2006
- The ARRL Antenna Book (15th edition), ARRL, 1988, ISBN 0-87259-206-5

### "Practical antenna" references

- UHF aerial positioning tests ([http://www.aerialsandtv.com/aerial\\_positioningtests.html](http://www.aerialsandtv.com/aerial_positioningtests.html))
- *Antenna Theory* antenna-theory.com (<http://www.antenna-theory.com>)
- *Patch Antenna: From Simulation to Realization* EM Talk ([http://www.emtalk.com/mwt\\_mpa.htm](http://www.emtalk.com/mwt_mpa.htm))
- *Why an Antenna Radiates* at ARRL (<http://www.arrl.org/tis/info/whyanradiates.html>)
- *Why Antennas Radiate*, Stuart G. Downs, WY6EE ([http://www.arrl.org/files/file/QEX\\_Binaries/0105downs.pdf](http://www.arrl.org/files/file/QEX_Binaries/0105downs.pdf)) (PDF)
- *Understanding electromagnetic fields and antenna radiation takes (almost) no math*, Ron Schmitt, EDN Magazine, March 2 2000 (<http://www.classicstla.com/download/emfields.pdf>) (PDF)
- Tests of FM/VHF receiving antennas. (<http://www.aerialsandtv.com/fmanddabradio.html#FMandDABaerialTests>)
- <http://www.tvantennasperth.com.au/Diyantennas.html> : "Antenna Gain"
- Antennas: Generalities, Principle of operation, As electronic component, Hertz Marconi and Other types Antennas etc etc (<http://www.ilmondodelletelecomunicazioni.it/english/antennas/index.htm>)

## Theory and simulations

- AN-SOF, " Antenna Simulation Software ([http://www.antennasoftware.com.ar/1\\_5\\_PRODUCTS.html](http://www.antennasoftware.com.ar/1_5_PRODUCTS.html))". Program system for the modeling of antennas and scatterers.
- <http://www.dipoleanimator.com>
- EM Talk, " Microstrip Patch Antenna ([http://www.emtalk.com/tut\\_1.htm](http://www.emtalk.com/tut_1.htm))", (Theory and simulation of microstrip patch antenna)
- " Online Calculations and Conversions (<http://www.jampro.com/index.php?page=technical-documents-and-calculators>)" Formulas for simulating and optimizing Antenna specs and placement
- " Microwave Antenna Design Calculator (<http://www.q-par.com/capabilities/software/microwave-antenna-design-calculator>)" Provides quick estimation of antenna size required for a given gain and frequency. 3 dB and 10 dB beamwidths are also derived; the calculator additionally gives the far-field range required for a given antenna.
- Sophocles J. Orfanidis, " Electromagnetic Waves and Antennas (<http://www.ece.rutgers.edu/~orfanidi/ewa/>)", Rutgers University (20 PDF Chaps. Basic theory, definitions and reference)
- Hans Lohninger, "Learning by Simulations: Physics: Coupled Radiators ([http://www.vias.org/simulations/simusoftware\\_twoaerials.html](http://www.vias.org/simulations/simusoftware_twoaerials.html))". *vias.org*, 2005. (ed. Interactive simulation of two coupled antennas)
- NEC Lab (<http://www.ingenierias.ugto.mx/profesores/sledesma/documentos/index.htm>) - NEC Lab is a tool that uses Numerical Electromagnetics Code and Artificial Intelligence to design and simulate antennas.
- Justin Smith " Aerials (<http://www.aerialsandtv.com/aerials.html>)". A.T.V (Aerials and Television), 2009. (ed. Article on the (basic) theory and use of FM, DAB & TV aerials)
- Antennas Research Group, " Virtual (Reality) Antennas (<http://www.antennas.gr>)". Democritus University of Thrace, 2005.
- "Support > Knowledgebase > RF Basics > Antennas / Cables > dBi vs. dBd detail (<http://www.maxstream.net/helpdesk/article-27>)". MaxStream, Inc., 2005. (ed. How to measure antenna gain) (New location: <http://www.digi.com/support/kbase/kbaseresultdet?id=2146> Note: to skip the registration form click the link below it)
- Yagis and Log Periodics, Astrosurf article. (<http://www.astrosurf.com/luxorion/qs1-antenna4.htm>)
- Raines, J. K., "Virtual Outer Conductor for Linear Antennas," *Microwave Journal*, Vol. 52, No. 1, January, 2009, pp. 76–86
- Tests of FM/VHF receiving antennas. (<http://www.aerialsandtv.com/fmanddabradio.html#FMandDABAerialTests>)

### Effect of ground references

- Electronic Radio and Engineering. F.E. Terman. McGraw-Hill
- Lectures on physics. Feynman, Leighton and Sands. Addison-Wesley
- Classical Electricity and Magnetism. W. Panofsky and M. Phillips. Addison-Wesley

## Patents and USPTO

- CLASS 343 (<http://www.uspto.gov/go/classification/uspc343/defs343.htm>), Communication: Radio Wave Antenna

## Further reading

- Antennas for Base Stations in Wireless Communications, edited by Zhi Ning Chen and Kwai-Man Luk, McGraw-Hill Companies, Inc, USA in May 2009



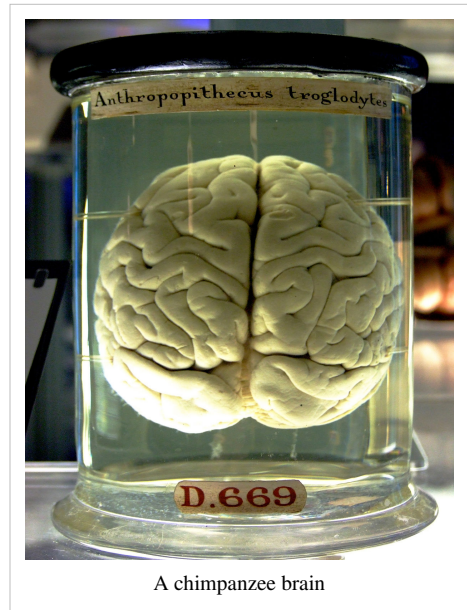
# Brain

The **brain** is the center of the nervous system in all vertebrate and most invertebrate animals—only a few invertebrates such as sponges, jellyfish, adult sea squirts and starfish do not have one, even if diffuse neural tissue is present. It is located in the head, usually close to the primary sensory organs for such senses as vision, hearing, balance, taste, and smell. The brain of a vertebrate is the most complex organ of its body. In a typical human the cerebral cortex (the largest part) is estimated to contain 15–33 billion neurons,<sup>[1]</sup> each connected by synapses to several thousand other neurons. These neurons communicate with one another by means of long protoplasmic fibers called axons, which carry trains of signal pulses called action potentials to distant parts of the brain or body targeting specific recipient cells.

Physiologically, the function of the brain is to exert centralized control over the other organs of the body. The brain acts on the rest of the body both by generating patterns of muscle activity and by driving secretion of chemicals called hormones. This centralized control allows rapid and coordinated responses to changes in the environment. Some basic types of responsiveness such as reflexes can be mediated by the spinal cord or peripheral ganglia, but sophisticated purposeful control of behavior based on complex sensory input requires the information-integrating capabilities of a centralized brain.

From a philosophical point of view, what makes the brain special in comparison to other organs is that it forms the physical structure that generates the mind. As Hippocrates put it: "Men ought to know that from nothing else but the brain come joys, delights, laughter and sports, and sorrows, griefs, despondency, and lamentations."<sup>[2]</sup> Through much of history, the mind was thought to be separate from the brain. Even for present-day neuroscience, the mechanisms by which brain activity gives rise to consciousness and thought remain very challenging to understand: despite rapid scientific progress, much about how the brain works remains a mystery. The operations of individual brain cells are now understood in considerable detail, but the way they cooperate in ensembles of millions has been very difficult to decipher. The most promising approaches treat the brain as a biological computer, very different in mechanism from electronic computers, but similar in the sense that it acquires information from the surrounding world, stores it, and processes it in a variety of ways.

This article compares the properties of brains across the entire range of animal species, with the greatest attention to vertebrates. It deals with the human brain insofar as it shares the properties of other brains. The ways in which the human brain differs from other brains are covered in the human brain article. Several topics that might be covered here are instead covered there because much more can be said about them in a human context. The most important is brain disease and the effects of brain damage, covered in the human brain article because the most common diseases of the human brain either do not show up in other species, or else manifest themselves in different ways.

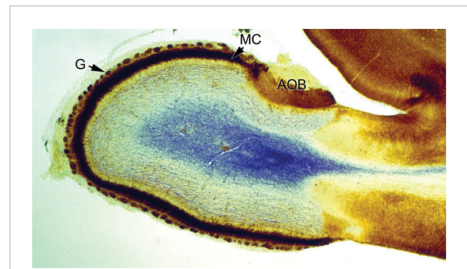


A chimpanzee brain

## Anatomy

The shape and size of the brains of different species vary greatly, and identifying common features is often difficult.<sup>[3]</sup> Nevertheless, there are a number of principles of brain architecture that apply across a wide range of species.<sup>[4]</sup> Some aspects of brain structure are common to almost the entire range of animals species;<sup>[5]</sup> others distinguish "advanced" brains from more primitive ones, or distinguish vertebrates from invertebrates.<sup>[3]</sup>

The simplest way to gain information about brain anatomy is by visual inspection, but many more sophisticated techniques have been developed. Brain tissue in its natural state is too soft to work with, but it can be hardened by immersion in alcohol or other fixatives, and then sliced apart for examination of the interior. Visually, the interior of the brain consists of areas of so-called grey matter, with a dark color, separated by areas of white matter, with a lighter color. Further information can be gained by staining slices of brain tissue with a variety of chemicals that bring out areas where specific types of molecules are present in high concentrations. It is also possible to examine the microstructure of brain tissue using a microscope, and to trace the pattern of connections from one brain area to another.<sup>[6]</sup>

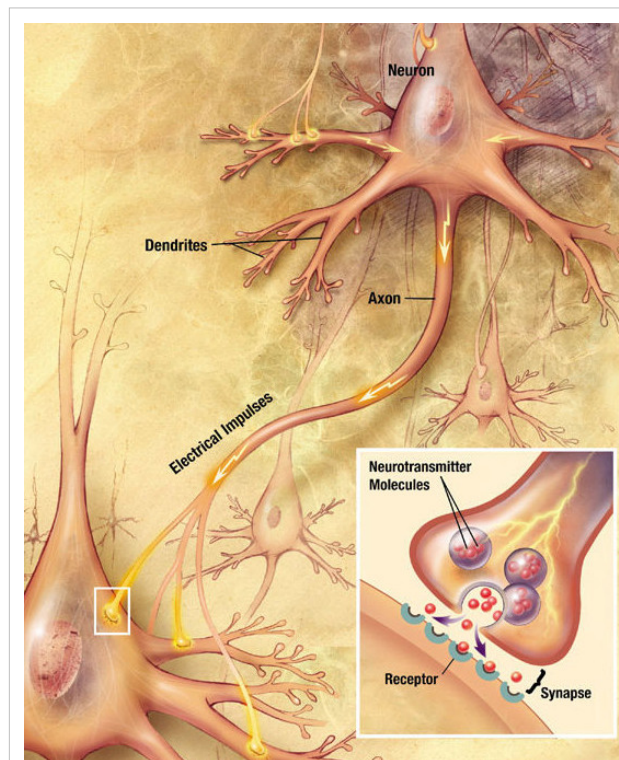


Cross section of the olfactory bulb of a rat, stained in two different ways at the same time: one stain shows neuron cell bodies, the other shows receptors for the neurotransmitter GABA.

## Cellular structure

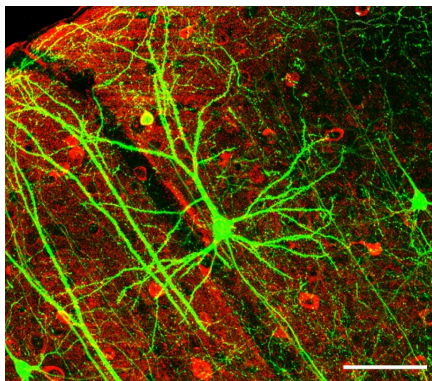
The brains of all species are composed primarily of two broad classes of cells: neurons and glial cells. Glial cells (also known as *glia* or *neuroglia*) come in several types, and perform a number of critical functions, including structural support, metabolic support, insulation, and guidance of development. Neurons, however, are usually considered the most important cells in the brain.<sup>[7]</sup>

The property that makes neurons unique is their ability to send signals to specific target cells over long distances.<sup>[8]</sup> They send these signals by means of an axon, which is a thin protoplasmic fiber that extends from the cell body and projects, usually with numerous branches, to other areas, sometimes nearby, sometimes in distant parts of the brain or body. The length of an axon can be extraordinary: for example, if a pyramidal cell of the cerebral cortex were magnified so that its cell body became the size of a human body, its axon, equally magnified, would become a cable a few centimeters in diameter, extending more than a kilometer.<sup>[9]</sup> These axons transmit signals in the form of electrochemical pulses called action potentials, which last less than a thousandth of a second and travel along the axon at speeds of 1–100 meters per second. Some neurons emit action potentials constantly, at rates of 10–100 per second, usually in irregular patterns; other neurons are quiet most of the time, but occasionally emit a burst of action potentials.<sup>[10]</sup>



Neurons generate electrical signals that travel along their axons. When a pulse of electricity reaches a junction called a synapse, it causes a neurotransmitter chemical to be released, which binds to receptors on other cells and thereby alters their electrical activity.

Axons transmit signals to other neurons by means of specialized junctions called synapses. A single axon may make as many as several thousand synaptic connections with other cells.<sup>[11]</sup> When an action potential, traveling along an axon, arrives at a synapse, it causes a chemical called a neurotransmitter to be released. The neurotransmitter binds to receptor molecules in the membrane of the target cell.<sup>[12]</sup>



Neurons often have extensive networks of dendrites, which receive synaptic connections. Shown is a pyramidal neuron from the hippocampus, stained for green fluorescent protein.

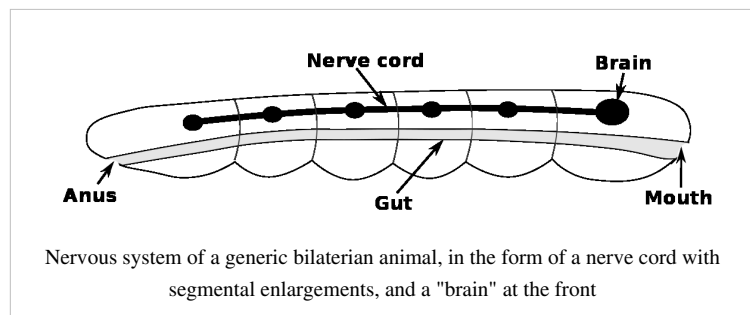
Synapses are the key functional elements of the brain.<sup>[13]</sup> The essential function of the brain is cell-to-cell communication, and synapses are the points at which communication occurs. The human brain has been estimated to contain approximately 100 trillion synapses;<sup>[14]</sup> even the brain of a fruit fly contains several million.<sup>[15]</sup> The functions of these synapses are very diverse: some are excitatory (excite the target cell); others are inhibitory; others work by activating second messenger systems that change the internal chemistry of their target cells in complex ways.<sup>[13]</sup> A large fraction of synapses are dynamically modifiable; that is, they are capable of changing strength in a way that is controlled by the patterns of signals that pass through them. It is widely believed that activity-dependent modification of synapses is the brain's primary mechanism for learning and memory.<sup>[13]</sup>

Most of the space in the brain is taken up by axons, which are often bundled together in what are called *nerve fiber tracts*. Many axons are wrapped in thick sheaths of a fatty substance called myelin, which serves to greatly increase the speed of signal propagation. Myelin is white, so parts of the brain filled exclusively with nerve fibers appear as light-colored white matter, in contrast to the darker-colored grey matter that marks areas with high densities of neuron cell bodies.<sup>[16]</sup>

## Evolution

### The generic bilaterian nervous system

Except for a few primitive types such as sponges (which have no nervous system<sup>[17]</sup>) and jellyfish (which have a nervous system consisting of a diffuse nerve net<sup>[17]</sup>), all living animals are bilaterians, meaning animals with a bilaterally symmetric body shape (that is, left and right sides that are approximate mirror images of each other).<sup>[18]</sup> All bilaterians are thought to have



Nervous system of a generic bilaterian animal, in the form of a nerve cord with segmental enlargements, and a "brain" at the front

descended from a common ancestor that appeared early in the Cambrian period, 550–600 million years ago, and it has been hypothesized that this common ancestor had the shape of a simple tubeworm with a segmented body.<sup>[18]</sup> At a schematic level, that basic worm-shape continues to be reflected in the body and nervous system architecture of all modern bilaterians, including vertebrates.<sup>[19]</sup> The fundamental bilateral body form is a tube with a hollow gut cavity running from the mouth to the anus, and a nerve cord with an enlargement (a ganglion) for each body segment, with an especially large ganglion at the front, called the brain. The brain is small and simple in some species, such as nematode worms; in other species, including vertebrates, it is the most complex organ in the body.<sup>[3]</sup> Some types of worms, such as leeches, also have an enlarged ganglion at the back end of the nerve cord, known as a "tail brain".<sup>[20]</sup>

There are a few types of existing bilaterians that lack a recognizable brain, including echinoderms, tunicates, and a group of primitive flatworms called Acoelomorpha. It has not been definitively established whether the existence of these brainless species indicates that the earliest bilaterians lacked a brain, or whether their ancestors evolved in a

way that led to the disappearance of a previously existing brain structure.<sup>[21]</sup>

## Invertebrates

This category includes arthropods, molluscs, and numerous types of worms. The diversity of invertebrate body plans is matched by an equal diversity in brain structures.<sup>[22]</sup>

Two groups of invertebrates have notably complex brains: arthropods (insects, crustaceans, arachnids, and others), and cephalopods (octopuses, squids, and similar molluscs).<sup>[23]</sup> The brains of arthropods and cephalopods arise from twin parallel nerve cords that extend through the body of the animal. Arthropods have a central brain with three divisions and large optical lobes behind each eye for visual processing.<sup>[23]</sup> Cephalopods such as the octopus and squid have the largest brains of any invertebrates.<sup>[24]</sup>

There are several invertebrate species whose brains have been studied intensively because they have properties that make them convenient for experimental work:

- Fruit flies (*Drosophila*), because of the large array of techniques available for studying their genetics, have been a natural subject for studying the role of genes in brain development.<sup>[25]</sup> In spite of the large evolutionary distance between insects and mammals, many aspects of *Drosophila* neurogenetics have turned out to be relevant to humans. The first biological clock genes, for example, were identified by examining *Drosophila* mutants that showed disrupted daily activity cycles.<sup>[26]</sup> A search in the genomes of vertebrates turned up a set of analogous genes, which were found to play similar roles in the mouse biological clock—and therefore almost certainly in the human biological clock as well.<sup>[27]</sup>
- The nematode worm *Caenorhabditis elegans*, like *Drosophila*, has been studied largely because of its importance in genetics.<sup>[28]</sup> In the early 1970s, Sydney Brenner chose it as a model system for studying the way that genes control development. One of the advantages of working with this worm is that the body plan is very stereotyped: the nervous system of the hermaphrodite morph contains exactly 302 neurons, always in the same places, making identical synaptic connections in every worm.<sup>[29]</sup> Brenner's team sliced worms into thousands of ultrathin sections and photographed every section under an electron microscope, then visually matched fibers from section to section, to map out every neuron and synapse in the entire body.<sup>[30]</sup> Nothing approaching this level of detail is available for any other organism, and the information has been used to enable a multitude of studies that would not have been possible without it.<sup>[31]</sup>
- The sea slug *Aplysia* was chosen by Nobel Prize-winning neurophysiologist Eric Kandel as a model for studying the cellular basis of learning and memory, because of the simplicity and accessibility of its nervous system, and it has been examined in hundreds of experiments.<sup>[32]</sup>

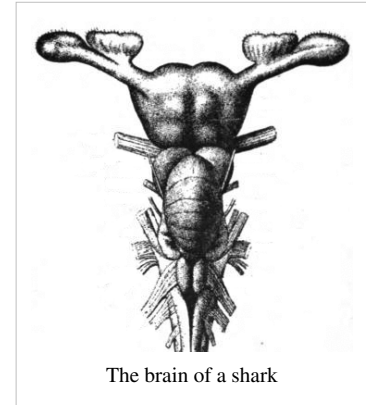


Fruit flies (*Drosophila*) have been extensively studied to gain insight into the role of genes in brain development.



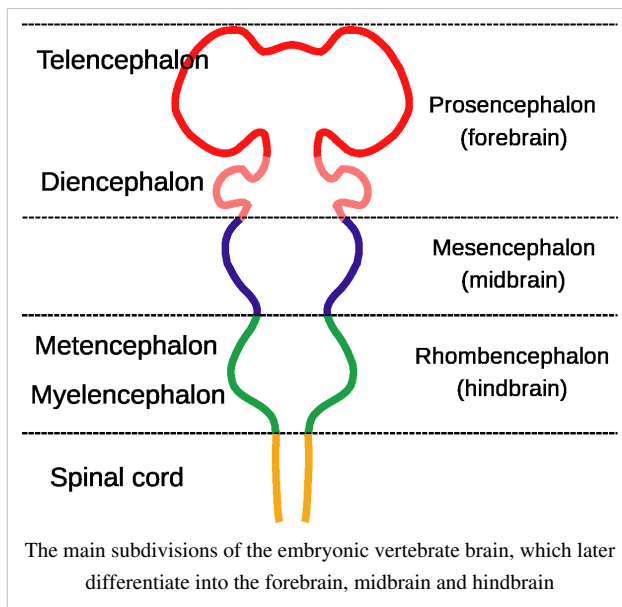
## Vertebrates

The first vertebrates appeared over 500 million years ago (Mya), during the Cambrian period, and may have resembled the modern hagfish in form.<sup>[33]</sup> Sharks appeared about 450 Mya, amphibians about 400 Mya, reptiles about 350 Mya, and mammals about 200 Mya. No modern species should be described as more "primitive" than others, strictly speaking, since each has an equally long evolutionary history—but the brains of modern hagfishes, lampreys, sharks, amphibians, reptiles, and mammals show a gradient of size and complexity that roughly follows the evolutionary sequence. All of these brains contain the same set of basic anatomical components, but many are rudimentary in the hagfish, whereas in mammals the foremost part (the telencephalon) is greatly elaborated and expanded.<sup>[34]</sup>



The brain of a shark

Brains are most simply compared in terms of their size. The relationship between brain size, body size and other variables has been studied across a wide range of vertebrate species. As a rule, brain size increases with body size, but not in a simple linear proportion. In general, smaller animals tend to have larger brains, measured as a fraction of body size: the animal with the largest brain-size-to-body-size ratio is the hummingbird. For mammals, the relationship between brain volume and body mass essentially follows a power law with an exponent of about 0.75.<sup>[35]</sup> This formula describes the central tendency, but every family of mammals departs from it to some degree, in a way that reflects in part the complexity of their behavior. For example, primates have brains 5 to 10 times larger than the formula predicts. Predators tend to have larger brains than their prey, relative to body size.<sup>[36]</sup>



All vertebrate brains share a common underlying form, which appears most clearly during early stages of embryonic development. In its earliest form, the brain appears as three swellings at the front end of the neural tube; these swellings eventually become the forebrain, midbrain, and hindbrain (the prosencephalon, mesencephalon, and rhombencephalon, respectively). At the earliest stages of brain development, the three areas are roughly equal in size. In many classes of vertebrates, such as fish and amphibians, the three parts remain similar in size in the adult, but in mammals the forebrain becomes much larger than the other parts, and the midbrain becomes very small.<sup>[37]</sup>

The brains of vertebrates are made of very soft tissue.<sup>[38]</sup> Living brain tissue is pinkish on the outside and mostly white on the inside, with subtle variations in

color. Vertebrate brains are surrounded by a system of connective tissue membranes called meninges that separate the skull from the brain. Blood vessels enter the central nervous system through holes in the meningeal layers. The cells in the blood vessel walls are joined tightly to one another, forming the so-called blood–brain barrier, which protects the brain from toxins that might enter through the bloodstream.<sup>[39]</sup>

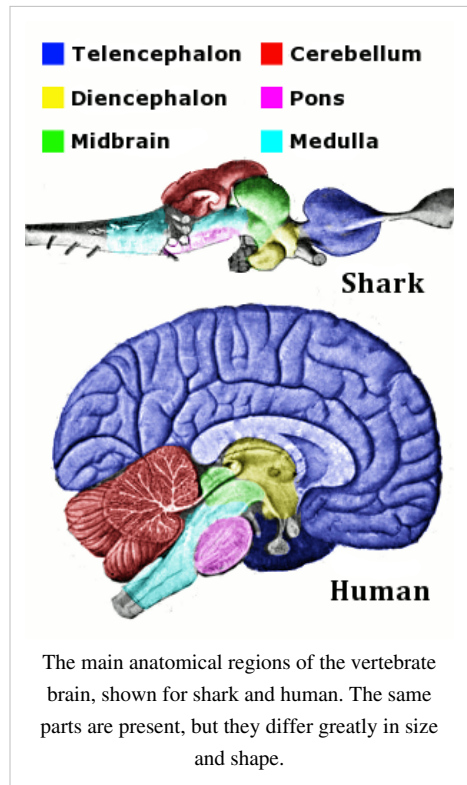
Neuroanatomists usually divide the vertebrate brain into six main regions: the telencephalon (cerebral hemispheres), diencephalon (thalamus and hypothalamus), mesencephalon (midbrain), cerebellum, pons, and medulla oblongata. Each of these areas has a complex internal structure. Some parts, such as the cerebral cortex and cerebellum, consist of layers that are folded or convoluted to fit within the available space. Other parts, such as the thalamus and hypothalamus, consist of clusters of many small nuclei. Thousands of distinguishable areas can be identified within

the vertebrate brain based on fine distinctions of neural structure, chemistry, and connectivity.<sup>[38]</sup>

Although the same basic components are present in all vertebrate brains, some branches of vertebrate evolution have led to substantial distortions of brain geometry, especially in the forebrain area. The brain of a shark shows the basic components in a straightforward way, but in teleost fishes (the great majority of existing fish species), the forebrain has become "everted", like a sock turned inside out. In birds, there are also major changes in forebrain structure.<sup>[40]</sup> These distortions can make it difficult to match brain components from one species with those of another species.<sup>[41]</sup>

Here is a list of some of the most important vertebrate brain components, along with a brief description of their functions as currently understood:

- The **medulla**, along with the spinal cord, contains many small nuclei involved in a wide variety of sensory and motor functions.<sup>[42]</sup>
- The **pons** lies in the brainstem directly above the medulla. Among other things, it contains nuclei that control sleep, respiration, swallowing, bladder function, equilibrium, eye movement, facial expressions, and posture.<sup>[43]</sup>
- The **hypothalamus** is a small region at the base of the forebrain, whose complexity and importance belies its size. It is composed of numerous small nuclei, each with distinct connections and neurochemistry. The hypothalamus regulates sleep and wake cycles, eating and drinking, hormone release, and many other critical biological functions.<sup>[44]</sup>
- The **thalamus** is another collection of nuclei with diverse functions. Some are involved in relaying information to and from the cerebral hemispheres. Others are involved in motivation. The subthalamic area (zona incerta) seems to contain action-generating systems for several types of "consummatory" behaviors, including eating, drinking, defecation, and copulation.<sup>[45]</sup>
- The **cerebellum** modulates the outputs of other brain systems to make them precise. Removal of the cerebellum does not prevent an animal from doing anything in particular, but it makes actions hesitant and clumsy. This precision is not built-in, but learned by trial and error. Learning how to ride a bicycle is an example of a type of neural plasticity that may take place largely within the cerebellum.<sup>[46]</sup>
- The **optic tectum** allows actions to be directed toward points in space, most commonly in response to visual input. In mammals it is usually referred to as the superior colliculus, and its best-studied function is to direct eye movements. It also directs reaching movements and other object-directed actions. It receives strong visual inputs, but also inputs from other senses that are useful in directing actions, such as auditory input in owls and input from the thermosensitive pit organs in snakes. In some fishes, such as lampreys, this region is the largest part of the brain.<sup>[47]</sup> The superior colliculus is part of the midbrain.
- The **pallium** is a layer of gray matter that lies on the surface of the forebrain. In reptiles and mammals, it is called the cerebral cortex. Multiple functions involve the pallium, including olfaction and spatial memory. In mammals, where it becomes so large as to dominate the brain, it takes over functions from many other brain areas. In many mammals, the cerebral cortex consists of folded bulges called gyri that create deep furrows or fissures called sulci. The folds increase the surface area of the cortex and therefore increase the amount of gray matter and the amount of information that can be processed.<sup>[48]</sup>
- The **hippocampus**, strictly speaking, is found only in mammals. However, the area it derives from, the medial pallium, has counterparts in all vertebrates. There is evidence that this part of the brain is involved in spatial memory and navigation in fishes, birds, reptiles, and mammals.<sup>[49]</sup>



- The **basal ganglia** are a group of interconnected structures in the forebrain. The primary function of the basal ganglia appears to be action selection: they send inhibitory signals to all parts of the brain that can generate motor behaviors, and in the right circumstances can release the inhibition, so that the action-generating systems are able to execute their actions. Reward and punishment exert their most important neural effects by altering connections within the basal ganglia.<sup>[50]</sup>
- The **olfactory bulb** is a special structure that processes olfactory sensory signals and sends its output to the olfactory part of the pallium. It is a major brain component in many vertebrates, but is greatly reduced in primates.<sup>[51]</sup>

## Mammals

The most obvious difference between the brains of mammals and other vertebrates is in terms of size. On average, a mammal has a brain roughly twice as large as that of a bird of the same body size, and ten times as large as that of a reptile of the same body size.<sup>[52]</sup>

Size, however, is not the only difference: there are also substantial differences in shape. The hindbrain and midbrain of mammals are generally similar to those of other vertebrates, but dramatic differences appear in the forebrain, which is greatly enlarged and also altered in structure.<sup>[53]</sup> The cerebral cortex is the part of the brain that most strongly distinguishes mammals. In non-mammalian vertebrates, the surface of the cerebrum is lined with a comparatively simple three-layered structure called the pallium. In mammals, the pallium evolves into a complex six-layered structure called neocortex or *isocortex*.<sup>[54]</sup> Several areas at the edge of the neocortex, including the hippocampus and amygdala, are also much more extensively developed in mammals than in other vertebrates.<sup>[53]</sup>

The elaboration of the cerebral cortex carries with it changes to other brain areas. The superior colliculus, which plays a major role in visual control of behavior in most vertebrates, shrinks to a small size in mammals, and many of its functions are taken over by visual areas of the cerebral cortex.<sup>[52]</sup> The cerebellum of mammals contains a large portion (the neocerebellum) dedicated to supporting the cerebral cortex, which has no counterpart in other vertebrates.<sup>[55]</sup>

## Primates

### Encephalization Quotient

Species	EQ <sup>[56]</sup>
Human	7.4–7.8
Chimpanzee	2.2–2.5
Rhesus monkey	2.1
Bottlenose dolphin	4.14 <sup>[57]</sup>
Elephant	1.13–2.36 <sup>[58]</sup>
Dog	1.2
Horse	0.9
Rat	0.4

The brains of humans and other primates contain the same structures as the brains of other mammals, but are generally larger in proportion to body size.<sup>[59]</sup> The most widely accepted way of comparing brain sizes across species is the so-called encephalization quotient (EQ), which takes into account the nonlinearity of the brain-to-body relationship.<sup>[56]</sup> Humans have an average EQ in the 7-to-8 range, while most other primates have an EQ in the 2-to-3 range. Dolphins have values higher than those of primates other than humans,<sup>[57]</sup> but nearly all other mammals have



EQ values that are substantially lower.

Most of the enlargement of the primate brain comes from a massive expansion of the cerebral cortex, especially the prefrontal cortex and the parts of the cortex involved in vision.<sup>[60]</sup> The visual processing network of primates includes at least 30 distinguishable brain areas, with a complex web of interconnections. It has been estimated that visual processing areas occupy more than half of the total surface of the primate neocortex.<sup>[61]</sup> The prefrontal cortex carries out functions that include planning, working memory, motivation, attention, and executive control. It takes up a much larger proportion of the brain for primates than for other species, and an especially large fraction of the human brain.<sup>[62]</sup>

## Physiology

The functions of the brain depend on the ability of neurons to transmit electrochemical signals to other cells, and their ability to respond appropriately to electrochemical signals received from other cells. The electrical properties of neurons are controlled by a wide variety of biochemical and metabolic processes, most notably the interactions between neurotransmitters and receptors that take place at synapses.<sup>[12]</sup>

## Neurotransmitters and receptors

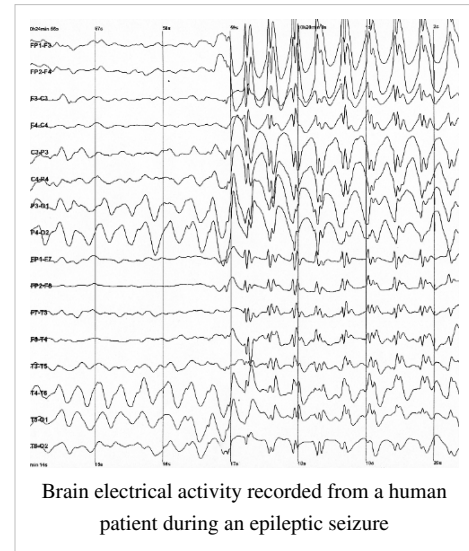
Neurotransmitters are chemicals that are released at synapses when an action potential activates them—neurotransmitters attach themselves to receptor molecules on the membrane of the synapse's target cell, and thereby alter the electrical or chemical properties of the receptor molecules. With few exceptions, each neuron in the brain releases the same chemical neurotransmitter, or combination of neurotransmitters, at all the synaptic connections it makes with other neurons; this rule is known as Dale's principle.<sup>[63]</sup> Thus, a neuron can be characterized by the neurotransmitters that it releases. The great majority of psychoactive drugs exert their effects by altering specific neurotransmitter systems. This applies to drugs such as marijuana, nicotine, heroin, cocaine, alcohol, fluoxetine, chlorpromazine, and many others.<sup>[64]</sup>

The two neurotransmitters that are used most widely in the vertebrate brain are glutamate, which almost always exerts excitatory effects on target neurons, and gamma-aminobutyric acid (GABA), which is almost always inhibitory. Neurons using these transmitters can be found in nearly every part of the brain.<sup>[65]</sup> Because of their ubiquity, drugs that act on glutamate or GABA tend to have broad and powerful effects. Some general anesthetics act by reducing the effects of glutamate; most tranquilizers exert their sedative effects by enhancing the effects of GABA.<sup>[66]</sup>

There are dozens of other chemical neurotransmitters that are used in more limited areas of the brain, often areas dedicated to a particular function. Serotonin, for example—the primary target of antidepressant drugs and many dietary aids—comes exclusively from a small brainstem area called the Raphe nuclei.<sup>[67]</sup> Norepinephrine, which is involved in arousal, comes exclusively from a nearby small area called the locus coeruleus.<sup>[68]</sup> Other neurotransmitters such as acetylcholine and dopamine have multiple sources in the brain, but are not as ubiquitously distributed as glutamate and GABA.<sup>[69]</sup>

## Electrical activity

As a side effect of the electrochemical processes used by neurons for signaling, brain tissue generates electric fields when it is active. When large numbers of neurons show synchronized activity, the electric fields that they generate can be large enough to detect outside the skull, using electroencephalography (EEG)<sup>[70]</sup> or magnetoencephalography (MEG). EEG recordings, along with recordings made from electrodes implanted inside the brains of animals such as rats, show that the brain of a living animal is constantly active, even during sleep.<sup>[71]</sup> Each part of the brain shows a mixture of rhythmic and nonrhythmic activity, which may vary according to behavioral state. In mammals, the cerebral cortex tends to show large slow delta waves during sleep, faster alpha waves when the animal is awake but inattentive, and chaotic-looking irregular activity when the animal is actively engaged in a task. During an epileptic seizure, the brain's inhibitory control mechanisms fail to function and electrical activity rises to pathological levels, producing EEG traces that show large wave and spike patterns not seen in a healthy brain. Relating these population-level patterns to the computational functions of individual neurons is a major focus of current research in neurophysiology.<sup>[71]</sup>



## Metabolism

All vertebrates have a blood–brain barrier that allows metabolism inside the brain to operate differently from metabolism in other parts of the body. Glial cells play a major role in brain metabolism, by controlling the chemical composition of the fluid that surrounds neurons, including levels of ions and nutrients.<sup>[72]</sup>

Brain tissue consumes a large amount of energy in proportion to its volume, so large brains place severe metabolic demands on animals. The need to limit body weight in order, for example, to fly, has apparently led to selection for a reduction of brain size in some species, such as bats.<sup>[73]</sup> Most of the brain's energy consumption goes into sustaining the electric charge (membrane potential) of neurons.<sup>[72]</sup> Most vertebrate species devote between 2% and 8% of basal metabolism to the brain. In primates, however, the fraction is much higher—in humans it rises to 20–25%.<sup>[74]</sup> The energy consumption of the brain does not vary greatly over time, but active regions of the cerebral cortex consume somewhat more energy than inactive regions; this forms the basis for the functional brain imaging methods PET, fMRI,<sup>[75]</sup> and NIRS.<sup>[76]</sup> In humans and many other species, the brain gets most of its energy from oxygen-dependent metabolism of glucose (i.e., blood sugar).<sup>[72]</sup> In some species, though, alternative sources of energy may be used, including lactate, ketones, amino acids, glycogen, and possibly lipids.<sup>[77]</sup>

## Functions

From an evolutionary-biological perspective, the function of the brain is to provide coherent control over the actions of an animal. A centralized brain allows groups of muscles to be co-activated in complex patterns; it also allows stimuli impinging on one part of the body to evoke responses in other parts, and it can prevent different parts of the body from acting at cross-purposes to each other.<sup>[78]</sup>

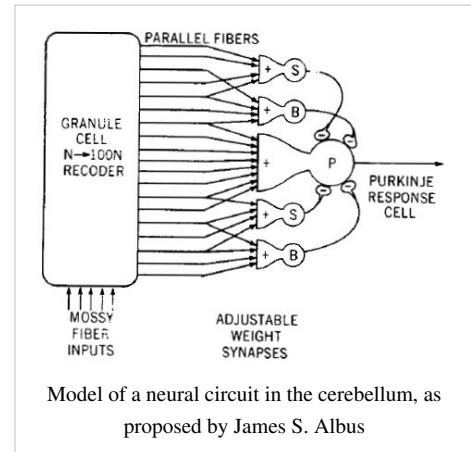
To generate purposeful and unified action, the brain first brings information from sense organs together at a central location. It then processes this raw data to extract information about the structure of the environment. Next it combines the processed sensory information with information about the current needs of an animal and with memory of past circumstances. Finally, on the basis of the results, it generates motor response patterns that are suited to maximize the welfare of the animal. These signal-processing tasks require intricate interplay between a variety of

functional subsystems.<sup>[78]</sup>

## Information processing

The invention of electronic computers in the 1940s, along with the development of mathematical information theory, led to a realization that brains can potentially be understood as information processing systems. This concept formed the basis of the field of cybernetics, and eventually gave rise to the field now known as computational neuroscience.<sup>[79]</sup> The earliest attempts at cybernetics were somewhat crude in that they treated the brain as essentially a digital computer in disguise, as for example in John von Neumann's 1958 book, *The Computer and the Brain*.<sup>[80]</sup> Over the years, though, accumulating information about the electrical responses of brain cells recorded from behaving animals has steadily moved theoretical concepts in the direction of increasing realism.<sup>[79]</sup>

The essence of the information processing approach is to try to understand brain function in terms of information flow and implementation of algorithms.<sup>[79]</sup> One of the most influential early contributions was a 1959 paper titled *What the frog's eye tells the frog's brain*: the paper examined the visual responses of neurons in the retina and optic tectum of frogs, and came to the conclusion that some neurons in the tectum of the frog are wired to combine elementary responses in a way that makes them function as "bug perceivers".<sup>[81]</sup> A few years later David Hubel and Torsten Wiesel discovered cells in the primary visual cortex of monkeys that become active when sharp edges move across specific points in the field of view—a discovery that eventually brought them a Nobel Prize.<sup>[82]</sup> Follow-up studies in higher-order visual areas found cells that detect binocular disparity, color, movement, and aspects of shape, with areas located at increasing distances from the primary visual cortex showing increasingly complex responses.<sup>[83]</sup> Other investigations of brain areas unrelated to vision have revealed cells with a wide variety of response correlates, some related to memory, some to abstract types of cognition such as space.<sup>[84]</sup>

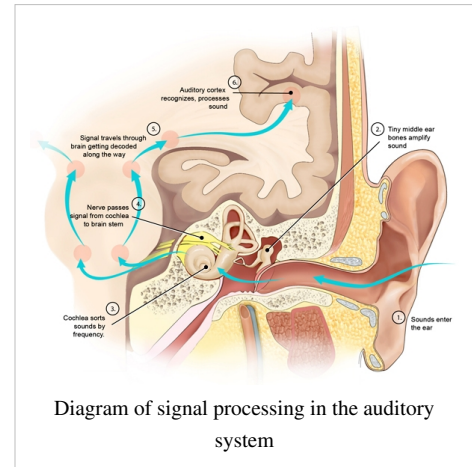


Model of a neural circuit in the cerebellum, as proposed by James S. Albus

Theorists have worked to understand these response patterns by constructing mathematical models of neurons and neural networks, which can be simulated using computers.<sup>[79]</sup> Some useful models are abstract, focusing on the conceptual structure of neural algorithms rather than the details of how they are implemented in the brain; other models attempt to incorporate data about the biophysical properties of real neurons.<sup>[85]</sup> No model on any level is yet considered to be a fully valid description of brain function, though. The essential difficulty is that sophisticated computation by neural networks requires distributed processing in which hundreds or thousands of neurons work cooperatively—current methods of brain activity recording are only capable of isolating action potentials from a few dozen neurons at a time.<sup>[86]</sup>

## Perception

One of the primary functions of a brain is to extract biologically relevant information from sensory inputs. The human brain is provided with information about light, sound, the chemical composition of the atmosphere, temperature, head orientation, limb position, the chemical composition of the bloodstream, and more. In other animals additional senses may be present, such as the infrared heat-sense of snakes, the magnetic field sense of some birds, or the electric field sense of some types of fish. Moreover, other animals may develop existing sensory systems in new ways, such as the adaptation by bats of the auditory sense into a form of sonar. One way or another, all of these sensory modalities are initially detected by specialized sensors that project signals into the brain.<sup>[87]</sup>



Each sensory system begins with specialized receptor cells, such as light-receptive neurons in the retina of the eye, vibration-sensitive neurons in the cochlea of the ear, or pressure-sensitive neurons in the skin. The axons of sensory receptor cells travel into the spinal cord or brain, where they transmit their signals to a first-order sensory nucleus dedicated to one specific sensory modality. This primary sensory nucleus sends information to higher-order sensory areas that are dedicated to the same modality. Eventually, via a way-station in the thalamus, the signals are sent to the cerebral cortex, where they are processed to extract biologically relevant features, and integrated with signals coming from other sensory systems.<sup>[87]</sup>

## Motor control

Motor systems are areas of the brain that are directly or indirectly involved in producing body movements, that is, in activating muscles. Except for the muscles that control the eye, which are driven by nuclei in the midbrain, all the voluntary muscles in the body are directly innervated by motor neurons in the spinal cord and hindbrain.<sup>[88]</sup> Spinal motor neurons are controlled both by neural circuits intrinsic to the spinal cord, and by inputs that descend from the brain. The intrinsic spinal circuits implement many reflex responses, and contain pattern generators for rhythmic movements such as walking or swimming. The descending connections from the brain allow for more sophisticated control.<sup>[89]</sup>

The brain contains several motor areas that project directly to the spinal cord. At the lowest level are motor areas in the medulla and pons, which control stereotyped movements such as walking, breathing, or swallowing. At a higher level are areas in the midbrain, such as the red nucleus, which is responsible for coordinating movements of the arms and legs. At a higher level yet is the primary motor cortex, a strip of tissue located at the posterior edge of the frontal lobe. The primary motor cortex sends projections to the subcortical motor areas, but also sends a massive projection directly to the spinal cord, through the pyramidal tract. This direct corticospinal projection allows for precise voluntary control of the fine details of movements. Other motor-related brain areas exert secondary effects by projecting to the primary motor areas. Among the most important secondary areas are the premotor cortex, basal ganglia, and cerebellum.<sup>[90]</sup>

### Major areas involved in controlling movement

Area	Location	Function
Ventral horn	Spinal cord	Contains motor neurons that directly activate muscles <sup>[91]</sup>
Oculomotor nuclei	Midbrain	Contains motor neurons that directly activate the eye muscles <sup>[92]</sup>
Cerebellum	Hindbrain	Calibrates precision and timing of movements <sup>[46]</sup>
Basal ganglia	Forebrain	Action selection on the basis of motivation <sup>[93]</sup>
Motor cortex	Frontal lobe	Direct cortical activation of spinal motor circuits
Premotor cortex	Frontal lobe	Groups elementary movements into coordinated patterns <sup>[94]</sup>
Supplementary motor area	Frontal lobe	Sequences movements into temporal patterns <sup>[95]</sup>
Prefrontal cortex	Frontal lobe	Planning and other executive functions <sup>[96]</sup>

In addition to all of the above, the brain and spinal cord contain extensive circuitry to control the autonomic nervous system, which works by secreting hormones and by modulating the "smooth" muscles of the gut.<sup>[97]</sup> The autonomic nervous system affects heart rate, digestion, respiration rate, salivation, perspiration, urination, and sexual arousal, and several other processes. Most of its functions are not under direct voluntary control.

### Arousal

Perhaps the most obvious aspect of the behavior of any animal is the daily cycle between sleeping and waking. Arousal and alertness are also modulated on a finer time scale, though, by an extensive network of brain areas.<sup>[98]</sup>

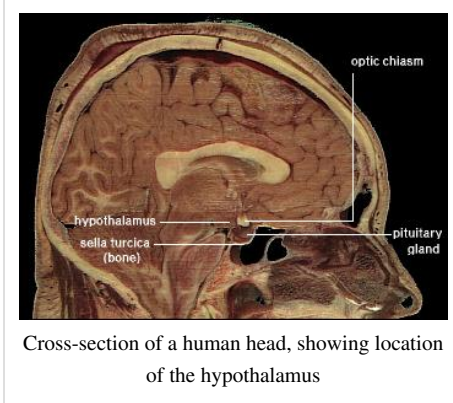
A key component of the arousal system is the suprachiasmatic nucleus (SCN), a tiny part of the hypothalamus located directly above the point at which the optic nerves from the two eyes cross. The SCN contains the body's central biological clock. Neurons there show activity levels that rise and fall with a period of about 24 hours, circadian rhythms: these activity fluctuations are driven by rhythmic changes in expression of a set of "clock genes". The SCN continues to keep time even if it is excised from the brain and placed in a dish of warm nutrient solution, but it ordinarily receives input from the optic nerves, through the retinohypothalamic tract (RHT), that allows daily light-dark cycles to calibrate the clock.<sup>[99]</sup>

The SCN projects to a set of areas in the hypothalamus, brainstem, and midbrain that are involved in implementing sleep-wake cycles. An important component of the system is the reticular formation, a group of neuron-clusters scattered diffusely through the core of the lower brain. Reticular neurons send signals to the thalamus, which in turn sends activity-level-controlling signals to every part of the cortex. Damage to the reticular formation can produce a permanent state of coma.<sup>[98]</sup>

Sleep involves great changes in brain activity.<sup>[100]</sup> Until the 1950s it was generally believed that the brain essentially shuts off during sleep,<sup>[101]</sup> but this is now known to be far from true; activity continues, but patterns become very different. There are two types of sleep: *REM sleep* (with dreaming) and *NREM* (non-REM, usually without dreaming) sleep, which repeat in slightly varying patterns throughout a sleep episode. Three broad types of distinct brain activity patterns can be measured: REM, light NREM and deep NREM. During deep NREM sleep, also called slow wave sleep, activity in the cortex takes the form of large synchronized waves, whereas in the waking state it is noisy and desynchronized. Levels of the neurotransmitters norepinephrine and serotonin drop during slow wave sleep, and fall almost to zero during REM sleep; levels of acetylcholine show the reverse pattern.<sup>[100]</sup>

## Homeostasis

For any animal, survival requires maintaining a variety of parameters of bodily state within a limited range of variation: these include temperature, water content, salt concentration in the bloodstream, blood glucose levels, blood oxygen level, and others.<sup>[102]</sup> The ability of an animal to regulate the internal environment of its body—the *milieu intérieur*, as pioneering physiologist Claude Bernard called it—is known as homeostasis (Greek for "standing still").<sup>[103]</sup> Maintaining homeostasis is a crucial function of the brain. The basic principle that underlies homeostasis is negative feedback: any time a parameter diverges from its set-point, sensors generate an error signal that evokes a response that causes the parameter to shift back toward its optimum value.<sup>[102]</sup> (This principle is widely used in engineering, for example in the control of temperature using a thermostat.)

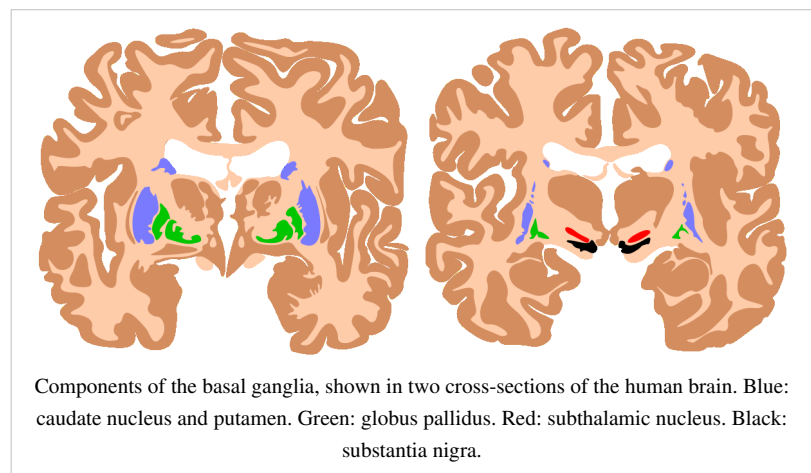


Cross-section of a human head, showing location of the hypothalamus

In vertebrates, the part of the brain that plays the greatest role is the hypothalamus, a small region at the base of the forebrain whose size does not reflect its complexity or the importance of its function.<sup>[102]</sup> The hypothalamus is a collection of small nuclei, most of which are involved in basic biological functions. Some of these functions relate to arousal or to social interactions such as sexuality, aggression, or maternal behaviors; but many of them relate to homeostasis. Several hypothalamic nuclei receive input from sensors located in the lining of blood vessels, conveying information about temperature, sodium level, glucose level, blood oxygen level, and other parameters. These hypothalamic nuclei send output signals to motor areas that can generate actions to rectify deficiencies. Some of the outputs also go to the pituitary gland, a tiny gland attached to the brain directly underneath the hypothalamus. The pituitary gland secretes hormones into the bloodstream, where they circulate throughout the body and induce changes in cellular activity.<sup>[104]</sup>

## Motivation

According to evolutionary theory, all species are genetically programmed to act as though they have a goal of surviving and propagating offspring. At the level of an individual animal, this overarching goal of genetic fitness translates into a set of specific survival-promoting behaviors, such as seeking food, water, shelter, and a mate.<sup>[105]</sup> The motivational system in the brain monitors the current state of satisfaction of these goals, and activates behaviors to meet any needs



Components of the basal ganglia, shown in two cross-sections of the human brain. Blue: caudate nucleus and putamen. Green: globus pallidus. Red: subthalamic nucleus. Black: substantia nigra.

that arise. The motivational system works largely by a reward–punishment mechanism. When a particular behavior is followed by favorable consequences, the reward mechanism in the brain is activated, which induces structural changes inside the brain that cause the same behavior to be repeated later, whenever a similar situation arises. Conversely, when a behavior is followed by unfavorable consequences, the brain's punishment mechanism is activated, inducing structural changes that cause the behavior to be suppressed when similar situations arise in the future.<sup>[106]</sup>

Every type of animal brain that has been studied uses a reward–punishment mechanism: even worms and insects can alter their behavior to seek food sources or to avoid dangers.<sup>[107]</sup> In vertebrates, the reward-punishment system is implemented by a specific set of brain structures, at the heart of which lie the basal ganglia, a set of interconnected areas at the base of the forebrain.<sup>[50]</sup> There is substantial evidence that the basal ganglia are the central site at which decisions are made: the basal ganglia exert a sustained inhibitory control over most of the motor systems in the brain; when this inhibition is released, a motor system is permitted to execute the action it is programmed to carry out. Rewards and punishments function by altering the relationship between the inputs that the basal ganglia receive and the decision-signals that are emitted. The reward mechanism is better understood than the punishment mechanism, because its role in drug abuse has caused it to be studied very intensively. Research has shown that the neurotransmitter dopamine plays a central role: addictive drugs such as cocaine, amphetamine, and nicotine either cause dopamine levels to rise or cause the effects of dopamine inside the brain to be enhanced.<sup>[108]</sup>

## Learning and memory

Almost all animals are capable of modifying their behavior as a result of experience—even the most primitive types of worms. Because behavior is driven by brain activity, changes in behavior must somehow correspond to changes inside the brain. Theorists dating back to Santiago Ramón y Cajal argued that the most plausible explanation is that learning and memory are expressed as changes in the synaptic connections between neurons.<sup>[109]</sup> Until 1970, however, experimental evidence to support the synaptic plasticity hypothesis was lacking. In 1971 Tim Bliss and Terje Lømo published a paper on a phenomenon now called long-term potentiation: the paper showed clear evidence of activity-induced synaptic changes that lasted for at least several days.<sup>[110]</sup> Since then technical advances have made these sorts of experiments much easier to carry out, and thousands of studies have been made that have clarified the mechanism of synaptic change, and uncovered other types of activity-driven synaptic change in a variety of brain areas, including the cerebral cortex, hippocampus, basal ganglia, and cerebellum.<sup>[111]</sup>

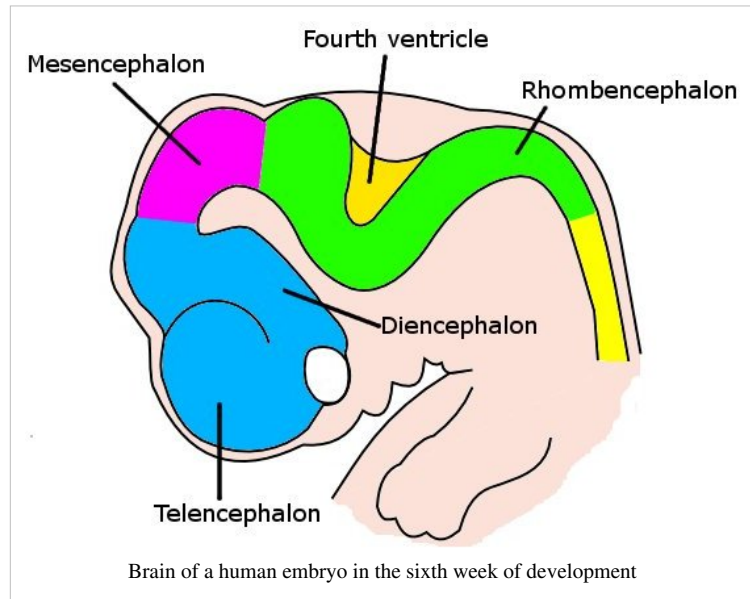
Neuroscientists currently distinguish several types of learning and memory that are implemented by the brain in distinct ways:

- **Working memory** is the ability of the brain to maintain a temporary representation of information about the task that an animal is currently engaged in. This sort of dynamic memory is thought to be mediated by the formation of cell assemblies—groups of activated neurons that maintain their activity by constantly stimulating one another.<sup>[112]</sup>
- **Episodic memory** is the ability to remember the details of specific events. This sort of memory can last for a lifetime. Much evidence implicates the hippocampus in playing a crucial role: people with severe damage to the hippocampus sometimes show amnesia, that is, inability to form new long-lasting episodic memories.<sup>[113]</sup>
- **Semantic memory** is the ability to learn facts and relationships. This sort of memory is probably stored largely in the cerebral cortex, mediated by changes in connections between cells that represent specific types of information.<sup>[114]</sup>
- **Instrumental learning** is the ability for rewards and punishments to modify behavior. It is implemented by a network of brain areas centered on the basal ganglia.<sup>[115]</sup>
- **Motor learning** is the ability to refine patterns of body movement by practicing, or more generally by repetition. A number of brain areas are involved, including the premotor cortex, basal ganglia, and especially the cerebellum, which functions as a large memory bank for microadjustments of the parameters of movement.<sup>[116]</sup>



## Development

The brain does not simply grow, but rather develops in an intricately orchestrated sequence of stages.<sup>[117]</sup> It changes in shape from a simple swelling at the front of the nerve cord in the earliest embryonic stages, to a complex array of areas and connections. Neurons are created in special zones that contain stem cells, and then migrate through the tissue to reach their ultimate locations. Once neurons have positioned themselves, their axons sprout and navigate through the brain, branching and extending as they go, until the tips reach their targets and form synaptic connections. In a number of parts of the nervous system, neurons and synapses are produced in excessive numbers during the early stages, and then the unneeded ones are pruned away.<sup>[118]</sup>



For vertebrates, the early stages of neural development are similar across all species.<sup>[117]</sup> As the embryo transforms from a round blob of cells into a wormlike structure, a narrow strip of ectoderm running along the midline of the back is induced to become the neural plate, the precursor of the nervous system. The neural plate folds inward to form the neural groove, and then the lips that line the groove merge to enclose the neural tube, a hollow cord of cells with a fluid-filled ventricle at the center. At the front end, the ventricles and cord swell to form three vesicles that are the precursors of the forebrain, midbrain, and hindbrain. At the next stage, the forebrain splits into two vesicles called the telencephalon (which will contain the cerebral cortex, basal ganglia, and related structures) and the diencephalon (which will contain the thalamus and hypothalamus). At about the same time, the hindbrain splits into the metencephalon (which will contain the cerebellum and pons) and the myelencephalon (which will contain the medulla oblongata). Each of these areas contains proliferative zones where neurons and glial cells are generated; the resulting cells then migrate, sometimes for long distances, to their final positions.<sup>[117]</sup>

Once a neuron is in place, it extends dendrites and an axon into the area around it. Axons, because they commonly extend a great distance from the cell body and need to reach specific targets, grow in a particularly complex way. The tip of a growing axon consists of a blob of protoplasm called a growth cone, studded with chemical receptors. These receptors sense the local environment, causing the growth cone to be attracted or repelled by various cellular elements, and thus to be pulled in a particular direction at each point along its path. The result of this pathfinding process is that the growth cone navigates through the brain until it reaches its destination area, where other chemical cues cause it to begin generating synapses. Considering the entire brain, thousands of genes create products that influence axonal pathfinding.<sup>[119]</sup>

The synaptic network that finally emerges is only partly determined by genes, though. In many parts of the brain, axons initially "overgrow", and then are "pruned" by mechanisms that depend on neural activity.<sup>[120]</sup> In the projection from the eye to the midbrain, for example, the structure in the adult contains a very precise mapping, connecting each point on the surface of the retina to a corresponding point in a midbrain layer. In the first stages of development, each axon from the retina is guided to the right general vicinity in the midbrain by chemical cues, but then branches very profusely and makes initial contact with a wide swath of midbrain neurons. The retina, before birth, contains special mechanisms that cause it to generate waves of activity that originate spontaneously at a random point and then propagate slowly across the retinal layer. These waves are useful because they cause neighboring neurons to be active at the same time; that is, they produce a neural activity pattern that contains

information about the spatial arrangement of the neurons. This information is exploited in the midbrain by a mechanism that causes synapses to weaken, and eventually vanish, if activity in an axon is not followed by activity of the target cell. The result of this sophisticated process is a gradual tuning and tightening of the map, leaving it finally in its precise adult form.<sup>[121]</sup>

Similar things happen in other brain areas: an initial synaptic matrix is generated as a result of genetically determined chemical guidance, but then gradually refined by activity-dependent mechanisms, partly driven by internal dynamics, partly by external sensory inputs. In some cases, as with the retina-midbrain system, activity patterns depend on mechanisms that operate only in the developing brain, and apparently exist solely to guide development.<sup>[121]</sup>

In humans and many other mammals, new neurons are created mainly before birth, and the infant brain contains substantially more neurons than the adult brain.<sup>[122]</sup> There are, however, a few areas where new neurons continue to be generated throughout life. The two areas for which adult neurogenesis is well established are the olfactory bulb, which is involved in the sense of smell, and the dentate gyrus of the hippocampus, where there is evidence that the new neurons play a role in storing newly acquired memories. With these exceptions, however, the set of neurons that is present in early childhood is the set that is present for life. Glial cells are different: as with most types of cells in the body, they are generated throughout the lifespan.<sup>[123]</sup>

There has long been debate about whether the qualities of mind, personality, and intelligence can be attributed to heredity or to upbringing—this is the nature versus nurture controversy.<sup>[124]</sup> Although many details remain to be settled, neuroscience research has clearly shown that both factors are important. Genes determine the general form of the brain, and genes determine how the brain reacts to experience. Experience, however, is required to refine the matrix of synaptic connections, which in its developed form contains far more information than the genome does. In some respects, all that matters is the presence or absence of experience during critical periods of development.<sup>[125]</sup> In other respects, the quantity and quality of experience are important; for example, there is substantial evidence that animals raised in enriched environments have thicker cerebral cortices, indicating a higher density of synaptic connections, than animals whose levels of stimulation are restricted.<sup>[126]</sup>

## Research

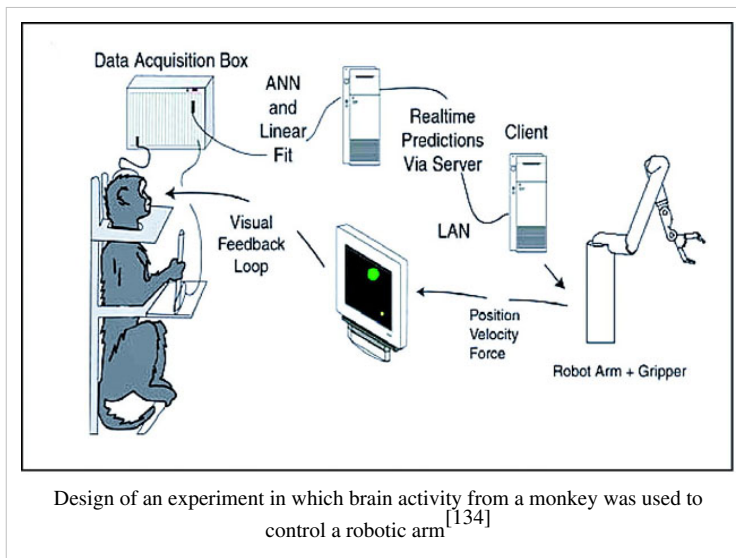
The field of neuroscience encompasses all approaches that seek to understand the brain and the rest of the nervous system.<sup>[127]</sup> Psychology seeks to understand mind and behavior, and neurology is the medical discipline that diagnoses and treats diseases of the nervous system. The brain is also the most important organ studied in psychiatry, the branch of medicine that works to study, prevent, and treat mental disorders.<sup>[128]</sup> Cognitive science seeks to unify neuroscience and psychology with other fields that concern themselves with the brain, such as computer science (artificial intelligence and similar fields) and philosophy.<sup>[129]</sup>

The oldest method of studying the brain is anatomical, and until the middle of the 20th century, much of the progress in neuroscience came from the development of better cell stains and better microscopes. Neuroanatomists study the large-scale structure of the brain as well as the microscopic structure of neurons and their components, especially synapses. Among other tools, they employ a plethora of stains that reveal neural structure, chemistry, and connectivity. In recent years, the development of immunostaining techniques has allowed investigation of neurons that express specific sets of genes. Also, *functional neuroanatomy* uses medical imaging techniques to correlate variations in human brain structure with differences in cognition or behavior.<sup>[130]</sup>



Human subject with EEG recording electrodes arranged around his head

Neurophysiologists study the chemical, pharmacological, and electrical properties of the brain: their primary tools are drugs and recording devices. Thousands of experimentally developed drugs affect the nervous system, some in highly specific ways. Recordings of brain activity can be made using electrodes, either glued to the scalp as in EEG studies, or implanted inside the brains of animals for extracellular recordings, which can detect action potentials generated by individual neurons.<sup>[131]</sup> Because the brain does not contain pain receptors, it is possible using these techniques to record brain activity from animals that are awake and behaving without causing distress. The same techniques have occasionally been used to study brain activity in human patients suffering from intractable epilepsy, in cases where there was a medical necessity to implant electrodes to localize the brain area responsible for epileptic seizures.<sup>[132]</sup> Functional imaging techniques such as functional magnetic resonance imaging are also used to study brain activity; these techniques have mainly been used with human subjects, because they require a conscious subject to remain motionless for long periods of time, but they have the great advantage of being noninvasive.<sup>[133]</sup>



Another approach to brain function is to examine the consequences of damage to specific brain areas. Even though it is protected by the skull and meninges, surrounded by cerebrospinal fluid, and isolated from the bloodstream by the blood–brain barrier, the delicate nature of the brain makes it vulnerable to numerous diseases and several types of damage. In humans, the effects of strokes and other types of brain damage have been a key source of information about brain function. Because there is no ability to experimentally control the nature of the damage, however, this information is often difficult to

interpret. In animal studies, most commonly involving rats, it is possible to use electrodes or locally injected chemicals to produce precise patterns of damage and then examine the consequences for behavior.<sup>[135]</sup>

Computational neuroscience encompasses two approaches: first, the use of computers to study the brain; second, the study of how brains perform computation. On one hand, it is possible to write a computer program to simulate the operation of a group of neurons by making use of systems of equations that describe their electrochemical activity; such simulations are known as *biologically realistic neural networks*. On the other hand, it is possible to study algorithms for neural computation by simulating, or mathematically analyzing, the operations of simplified "units" that have some of the properties of neurons but abstract out much of their biological complexity. The computational functions of the brain are studied both by computer scientists and neuroscientists.<sup>[136]</sup>

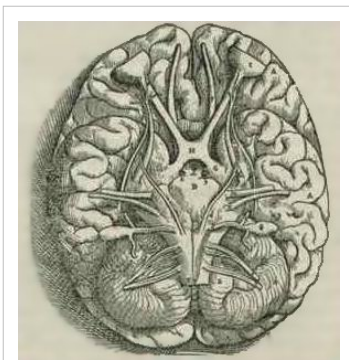
Recent years have seen increasing applications of genetic and genomic techniques to the study of the brain.<sup>[137]</sup> The most common subjects are mice, because of the availability of technical tools. It is now possible with relative ease to "knock out" or mutate a wide variety of genes, and then examine the effects on brain function. More sophisticated approaches are also being used: for example, using Cre-Lox recombination it is possible to activate or deactivate genes in specific parts of the brain, at specific times.<sup>[137]</sup>

## History

Early philosophers were divided as to whether the seat of the soul lies in the brain or heart. Aristotle favored the heart, and thought that the function of the brain was merely to cool the blood. Democritus, the inventor of the atomic theory of matter, argued for a three-part soul, with intellect in the head, emotion in the heart, and lust near the liver.<sup>[138]</sup> Hippocrates, the "father of medicine", came down unequivocally in favor of the brain. In his treatise on epilepsy he wrote:

Men ought to know that from nothing else but the brain come joys, delights, laughter and sports, and sorrows, griefs, despondency, and lamentations. ... And by the same organ we become mad and delirious, and fears and terrors assail us, some by night, and some by day, and dreams and untimely wanderings, and cares that are not suitable, and ignorance of present circumstances, desuetude, and unskillfulness. All these things we endure from the brain, when it is not healthy...

Hippocrates, *On the Sacred Disease*<sup>[2]</sup>



Andreas Vesalius' *Fabrica*, published in 1543, showing the base of the human brain, including optic chiasma, cerebellum, olfactory bulbs, etc.

The Roman physician Galen also argued for the importance of the brain, and theorized in some depth about how it might work. Galen traced out the anatomical relationships among brain, nerves, and muscles, demonstrating that all muscles in the body are connected to the brain through a branching network of nerves. He postulated that nerves activate muscles mechanically by carrying a mysterious substance he called *pneumata psychikon*, usually translated as "animal spirits".<sup>[138]</sup> Galen's ideas were widely known during the Middle Ages, but not much further progress came until the Renaissance, when detailed anatomical study resumed, combined with the theoretical speculations of René Descartes and those who followed him. Descartes, like Galen, thought of the nervous system in hydraulic terms. He believed that the highest cognitive functions are carried out by a non-physical *res cogitans*, but that the majority of behaviors of humans, and all behaviors of animals, could be explained mechanistically.<sup>[139]</sup>

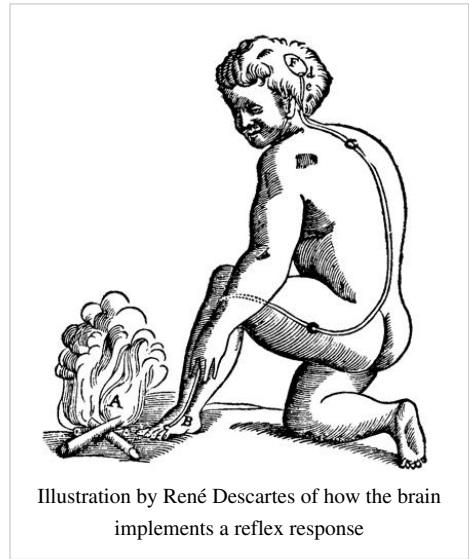


Illustration by René Descartes of how the brain implements a reflex response

The first real progress toward a modern understanding of nervous function, though, came from the investigations of Luigi Galvani, who discovered that a shock of static electricity applied to an exposed nerve of a dead frog could cause its leg to contract. Since that time, each major advance in understanding has followed more or less directly from the development of a new technique of investigation. Until the early years of the 20th century, the most important advances were derived from new methods for staining cells.<sup>[140]</sup> Particularly critical was the invention of the Golgi stain, which (when correctly used) stains only a small fraction of neurons, but stains them in their entirety, including cell body, dendrites, and axon. Without such a stain, brain tissue under a microscope appears as an impenetrable tangle of protoplasmic fibers, in which it is impossible to determine any structure. In the hands of Camillo Golgi, and especially of the Spanish neuroanatomist Santiago Ramón y Cajal, the new stain revealed hundreds of distinct types of neurons, each with its own unique dendritic structure and pattern of connectivity.<sup>[141]</sup>



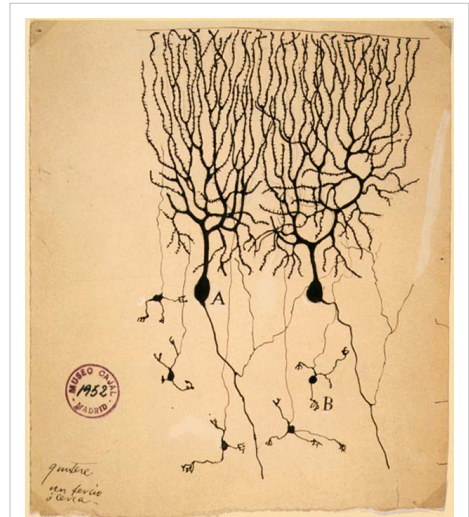
In the first half of the 20th century, advances in electronics enabled investigation of the electrical properties of nerve cells, culminating in work by Alan Hodgkin, Andrew Huxley, and others on the biophysics of the action potential, and the work of Bernard Katz and others on the electrochemistry of the synapse.<sup>[142]</sup> These studies complemented the anatomical picture with a conception of the brain as a dynamic entity. Reflecting the new understanding, in 1942 Charles Sherrington visualized the workings of the brain waking from sleep:

The great topmost sheet of the mass, that where hardly a light had twinkled or moved, becomes now a sparkling field of rhythmic flashing points with trains of traveling sparks hurrying hither and thither. The brain is waking and with it the mind is returning. It is as if the Milky Way entered upon some cosmic dance. Swiftly the head mass becomes an enchanted loom where millions of flashing shuttles weave a dissolving pattern, always a meaningful pattern though never an abiding one; a shifting harmony of subpatterns.

—Sherrington, 1942, *Man on his Nature*<sup>[143]</sup>

In the second half of the 20th century, developments in chemistry, electron microscopy, genetics, computer science, functional brain imaging, and other fields progressively opened new windows into brain structure and function. In the United States, the 1990s were officially designated as the "Decade of the Brain" to commemorate advances made in brain research, and to promote funding for such research.<sup>[144]</sup>

In the 21st century, these trends have continued, and several new approaches have come into prominence, including multielectrode recording, which allows the activity of many brain cells to be recorded all at the same time;<sup>[145]</sup> genetic engineering, which allows molecular components of the brain to be altered experimentally;<sup>[137]</sup> and genomics, which allows variations in brain structure to be correlated with variations in DNA properties.<sup>[146]</sup>



Drawing by Santiago Ramón y Cajal of two types of Golgi-stained neurons from the cerebellum of a pigeon

## References

- [1] Pelvig, DP; Pakkenberg, H; Stark, AK; Pakkenberg, B (2008). "Neocortical glial cell numbers in human brains". *Neurobiology of Aging* **29** (11): 1754–1762. doi:10.1016/j.neurobiolaging.2007.04.013. PMID 17544173.
- [2] Hippocrates (400 BCE). *On the Sacred Disease* (<http://classics.mit.edu/Hippocrates/sacred.html>). Francis Adams. .
- [3] Shepherd, GM (1994). *Neurobiology*. Oxford University Press. p. 3. ISBN 978-0-19-508843-4.
- [4] Sporns, O (2010). *Networks of the Brain* (<http://books.google.com/books?id=v1DBKE7-UrYC&pg=PA143>). MIT Press. p. 143. ISBN 978-0-262-01469-4. .
- [5] Başar, E (2010). *Brain-Body-Mind in the Nebulous Cartesian System: A Holistic Approach by Oscillations* (<http://books.google.com/books?id=NAbMH0-ux58C&pg=PA225>). Springer. p. 225. ISBN 978-1-4419-6134-1. .
- [6] Singh, I (2006). "A brief review of the techniques used in the study of neuroanatomy" (<http://books.google.com/books?id=bBMS013dmycC&pg=PA24>). *Textbook of human neuroanatomy*. Jaypee Brothers Publishers. p. 24. ISBN 978-81-8061-808-6. .
- [7] *Principles of Neural Science* p. 20
- [8] *Principles of Neural Science*, p. 21
- [9] Douglas, RJ; Martin, KA (2004). "Neuronal circuits of the neocortex". *Annual Review of Neuroscience* **27**: 419–451. doi:10.1146/annurev.neuro.27.070203.144152. PMID 15217339.
- [10] Barnett, MW; Larkman, PM (2007). "The action potential". *Practical Neurology* **7** (3): 192–197. PMID 17515599.
- [11] *Principles of Neural Science*, Ch.10, p. 175
- [12] *Principles of Neural Science*, Ch. 10
- [13] Shepherd, GM (2004). "Ch. 1: Introduction to synaptic circuits". *The Synaptic Organization of the Brain*. Oxford University Press US. ISBN 978-0-19-515956-1.
- [14] Williams, RW; Herrup, K (1988). "The control of neuron number". *Annual Review of Neuroscience* **11**: 423–453. doi:10.1146/annurev.ne.11.030188.002231. PMID 3284447.

- [15] Heisenberg, M (2003). "Mushroom body memoir: from maps to models". *Nature Reviews Neuroscience* **4** (4): 266–275. doi:10.1038/nrn1074. PMID 12671643.
- [16] *Principles of Neural Science*, Ch. 2
- [17] Jacobs, DK, Nakanishi N, Yuan D et al. (2007). "Evolution of sensory structures in basal metazoa" (<http://icb.oxfordjournals.org/cgi/content/full/47/5/712>). *Integrative & Comparative Biology* **47** (5): 712–723. doi:10.1093/icb/icm094. PMID 21669752. .
- [18] Balavoine, G (2003). "The segmented Urbilateria: A testable scenario" (<http://icb.oxfordjournals.org/cgi/content/full/43/1/137>). *Integrative & Comparative Biology* **43** (1): 137–147. doi:10.1093/icb/43.1.137. .
- [19] Schmidt-Rhaesa, A (2007). *The Evolution of Organ Systems*. Oxford University Press. p. 110. ISBN 978-0-19-856669-4.
- [20] Kristan Jr, WB; Calabrese, RL; Friesen, WO (2005). "Neuronal control of leech behavior". *Prog Neurobiology* **76** (5): 279–327. doi:10.1016/j.pneurobio.2005.09.004. PMID 16260077.
- [21] Mwinyi, A; Bailly, X; Bourlat, SJ; Jondelius, U; Littlewood, DT; Podsiadlowski, L (2010). "The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of *Symsagittifera roscoffensis*". *BMC Evolutionary Biology* **10**: 309. doi:10.1186/1471-2148-10-309. PMC 2973942. PMID 20942955.
- [22] Barnes, RD (1987). *Invertebrate Zoology* (5th ed.). Saunders College Pub.. p. 1. ISBN 978-0-03-008914-5.
- [23] Butler, AB (2000). "Chordate Evolution and the Origin of Craniates: An Old Brain in a New Head". *Anatomical Record* **261** (3): 111–125. doi:10.1002/1097-0185(20000615)261:3<111::AID-AR6>3.0.CO;2-F. PMID 10867629.
- [24] Bulloch, TH; Kutch, W (1995). "Are the main grades of brains different principally in numbers of connections or also in quality?" (<http://books.google.com/books?id=dW5e6FHOH-4C&pg=PA439>). In Breidbach O. *The nervous systems of invertebrates: an evolutionary and comparative approach*. Birkhäuser. p. 439. ISBN 978-3-7643-5076-5. .
- [25] "Flybrain: An online atlas and database of the *drosophila* nervous system" (<http://flybrain.neurobio.arizona.edu>). . Retrieved 2011-10-14.
- [26] Konopka, RJ; Benzer, S (1971). "Clock Mutants of *Drosophila melanogaster*". *Proc Nat Acad Sci U.S.A.* **68** (9): 2112–6. doi:10.1073/pnas.68.9.2112. PMC 389363. PMID 5002428.
- [27] Shin HS et al. (1985). "An unusual coding sequence from a *Drosophila* clock gene is conserved in vertebrates". *Nature* **317** (6036): 445–8. doi:10.1038/317445a0. PMID 2413365.
- [28] "WormBook: The online review of *C. elegans* biology" (<http://www.wormbook.org>). . Retrieved 2011-10-14.
- [29] Hobert, O (2005). *Specification of the nervous system* ([http://www.wormbook.org/chapters/www\\_specnervsys/specnervsys.html](http://www.wormbook.org/chapters/www_specnervsys/specnervsys.html)). In The *C. elegans* Research Community. "Wormbook". *WormBook*: 1–19. doi:10.1895/wormbook.1.12.1. PMID 18050401. .
- [30] White, JG; Southgate, E; Thomson, JN; Brenner, S (1986). "The Structure of the Nervous System of the Nematode *Caenorhabditis elegans*". *Phil. Trans. Roy. Soc. London (Biology)* **314** (1165): 1–340. doi:10.1098/rstb.1986.0056.
- [31] Hodgkin, J (2001). "*Caenorhabditis elegans*". In Brenner S, Miller JH. *Encyclopedia of Genetics*. Elsevier. pp. 251–256. ISBN 978-0-12-227080-2.
- [32] Kandel, ER (2007). *In Search of Memory: The Emergence of a New Science of Mind*. WW Norton. pp. 145–150. ISBN 978-0-393-32937-7.
- [33] Shu, DG; Morris, SC; Han, J; Zhang, Z-F; Yasui, K.; Janvier, P.; Chen, L.; Zhang, X.-L. et al. (2003). "Head and backbone of the Early Cambrian vertebrate *Haikouichthys*". *Nature* **421** (6922): 526–529. doi:10.1038/nature01264. PMID 12556891.
- [34] Striedter, GF (2005). "Ch. 3: Conservation in vertebrate brains". *Principles of Brain Evolution*. Sinauer Associates. ISBN 978-0-87893-820-9.
- [35] Armstrong, E (1983). "Relative brain size and metabolism in mammals". *Science* **220** (4603): 1302–1304. doi:10.1126/science.6407108. PMID 6407108.
- [36] Jerison, HJ (1973). *Evolution of the Brain and Intelligence*. Academic Press. pp. 55–74. ISBN 978-0-12-385250-2.
- [37] *Principles of Neural Science*, p. 1019
- [38] *Principles of Neural Science*, Ch. 17
- [39] Parent, A; Carpenter, MB (1995). "Ch. 1". *Carpenter's Human Neuroanatomy*. Williams & Wilkins. ISBN 978-0-683-06752-1.
- [40] Northcutt, RG (2008). "Forebrain evolution in bony fishes". *Brain Research Bulletin* **75** (2–4): 191–205. doi:10.1016/j.brainresbull.2007.10.058. PMID 18331871.
- [41] Reiner, A; Yamamoto, K; Karten, HJ (2005). "Organization and evolution of the avian forebrain". *The Anatomical Record Part A* **287** (1): 1080–1102. doi:10.1002/ar.a.20253. PMID 16206213.
- [42] *Principles of Neural Science*, Chs. 44, 45
- [43] Siegel, A; Sapru, HN (2010). *Essential Neuroscience*. Lippincott Williams & Wilkins. pp. 184–189. ISBN 978-0-7817-8383-5.
- [44] Swaab, DF; Boller, F; Aminoff, MJ (2003). *The Human Hypothalamus* (<http://books.google.com/?id=Js81Pr1PmaAC>). Elsevier. ISBN 978-0-444-51357-1. .
- [45] Jones, EG (1985). *The Thalamus* (<http://books.google.com/?id=WMxqAAAAMAAJ>). Plenum Press. ISBN 978-0-306-41856-3. .
- [46] *Principles of Neural Science*, Ch. 42
- [47] Saitoh, K; Ménard, A; Grillner, S (2007). "Tectal control of locomotion, steering, and eye movements in lamprey" (<http://jn.physiology.org/cgi/content/full/97/4/3093>). *Journal of Neurophysiology* **97** (4): 3093–3108. doi:10.1152/jn.00639.2006. PMID 17303814. .
- [48] Puelles, L (2001). "Thoughts on the development, structure and evolution of the mammalian and avian telencephalic pallium". *Phil. Trans. Roy. Soc. London B (Biological Sciences)* **356** (1414): 1583–1598. doi:10.1098/rstb.2001.0973. PMC 1088538. PMID 11604125.
- [49] Salas, C; Broglio, C; Rodríguez, F (2003). "Evolution of forebrain and spatial cognition in vertebrates: conservation across diversity". *Brain, Behavior and Evolution* **62** (2): 72–82. doi:10.1159/000072438. PMID 12937346.

- [50] Grillner, S et al. (2005). "Mechanisms for selection of basic motor programs—roles for the striatum and pallidum". *Trends in Neurosciences* **28** (7): 364–370. doi:10.1016/j.tins.2005.05.004. PMID 15935487.
- [51] Northcutt, RG (1981). "Evolution of the telencephalon in nonmammals". *Annual Review of Neuroscience* **4**: 301–350. doi:10.1146/annurev.ne.04.030181.001505. PMID 7013637.
- [52] Northcutt, RG (2002). "Understanding vertebrate brain evolution" (<http://icb.oxfordjournals.org/content/42/4/743.full>). *Integrative & Comparative Biology* **42** (4): 743–756. doi:10.1093/icb/42.4.743. PMID 21708771. .
- [53] Barton, RA; Harvey, PH (2000). "Mosaic evolution of brain structure in mammals". *Nature* **405** (6790): 1055–1058. doi:10.1038/35016580. PMID 10890446.
- [54] Aboitiz, F; Morales, D; Montiel, J (2003). "The evolutionary origin of the mammalian isocortex: Towards an integrated developmental and functional approach" (<http://www.bbsonline.org/Preprints/Aboitiz/Referees>). *Behavioral and Brain Sciences* **26** (5): 535–552. doi:10.1017/S0140525X03000128. PMID 15179935. .
- [55] Romer, AS; Parsons, TS (1977). *The Vertebrate Body*. Holt-Saunders International. p. 531. ISBN 0-03-910284-X.
- [56] Roth, G; Dicke, U (2005). "Evolution of the brain and Intelligence". *Trends in Cognitive Sciences* **9** (5): 250–257. doi:10.1016/j.tics.2005.03.005. PMID 15866152.
- [57] Marino, Lori (2004). "Cetacean Brain Evolution: Multiplication Generates Complexity" (<http://www.cogs.indiana.edu/spackled/2005readings/CetaceanBrainEvolution.pdf>) (PDF). *International Society for Comparative Psychology* (17): 1–16. . Retrieved 2010-08-29.
- [58] Shoshani, J; Kupsky, WJ; Marchant, GH (2006). "Elephant brain Part I: Gross morphology, functions, comparative anatomy, and evolution". *Brain Research Bulletin* **70** (2): 124–157. doi:10.1016/j.brainresbull.2006.03.016. PMID 16782503.
- [59] Finlay, BL; Darlington, RB; Nicastro, N (2001). "Developmental structure in brain evolution". *Behavioral and Brain Sciences* **24** (2): 263–308. doi:10.1017/S0140525X01003958. PMID 11530543.
- [60] Calvin, WH (1996). *How Brains Think* (<http://books.google.com/?id=z1r03ECL5A8C>). Basic Books. ISBN 978-0-465-07278-1. .
- [61] Sereno, MI; Dale, AM; Reppas, AM; Kwong, KK; Belliveau, JW; Brady, TJ; Rosen, BR; Tootell, RBH (1995). "Borders of multiple visual areas in human revealed by functional magnetic resonance imaging" (<http://www.cogsci.ucsd.edu/~sereno/papers/HumanRetin95.pdf>). *Science* (AAAS) **268** (5212): 889–893. doi:10.1126/science.7754376. PMID 7754376. .
- [62] Fuster, JM (2008). *The Prefrontal Cortex*. Elsevier. pp. 1–7. ISBN 978-0-12-373644-4.
- [63] *Principles of Neural Science*, Ch. 15
- [64] Cooper, JR; Bloom, FE; Roth, RH (2003). *The Biochemical Basis of Neuropharmacology* (<http://books.google.com/?id=e5I5gOwxVMkC>). Oxford University Press US. ISBN 978-0-19-514008-8. .
- [65] McGeer, PL; McGeer, EG (1989). "Chapter 15, Amino acid neurotransmitters". In G. Siegel et al. *Basic Neurochemistry*. Raven Press. pp. 311–332. ISBN 978-0-88167-343-2.
- [66] Foster, AC; Kemp, JA (2006). "Glutamate- and GABA-based CNS therapeutics". *Current Opinion in Pharmacology* **6** (1): 7–17. doi:10.1016/j.coph.2005.11.005. PMID 16377242.
- [67] Frazer, A; Hensler, JG (1999). "Understanding the neuroanatomical organization of serotonergic cells in the brain provides insight into the functions of this neurotransmitter". In Siegel, GJ. *Basic Neurochemistry* (Sixth ed.). Lippincott Williams & Wilkins. ISBN 0-397-51820-X.
- [68] Mehler, MF; Purpura, DP (2009). "Autism, fever, epigenetics and the locus coeruleus". *Brain Research Reviews* **59** (2): 388–392. doi:10.1016/j.brainresrev.2008.11.001. PMC 2668953. PMID 19059284.
- [69] Rang, HP (2003). *Pharmacology*. Churchill Livingstone. pp. 476–483. ISBN 0-443-07145-4.
- [70] Speckmann, E-J; Elger, CE (2004). "Introduction to the neurophysiological basis of the EEG and DC potentials". In Niedermeyer E, Lopes da Silva FH. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins. pp. 17–31. ISBN 0-7817-5126-8.
- [71] Buzsáki, G (2006). *Rhythms of the Brain*. Oxford University Press. ISBN 978-0-19-530106-9. OCLC 63279497.
- [72] Nieuwenhuys, R; Donkelaar, HJ; Nicholson, C (1998). *The Central Nervous System of Vertebrates, Volume 1*. Springer. pp. 11–14. ISBN 978-3-540-56013-5.
- [73] Safi, K; Seid, MA; Dechmann, DK (2005). "Bigger is not always better: when brains get smaller". *Biology Letters* **1** (3): 283–286. doi:10.1098/rsbl.2005.0333. PMC 1617168. PMID 17148188.
- [74] Mink, JW; Blumenshine, RJ; Adams, DB (1981). "Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis". *American Journal of Physiology* **241** (3): R203–212. PMID 7282965.
- [75] Raichle, M; Gusnard, DA (2002). "Appraising the brain's energy budget". *Proc. Nat. Acad. Sci. U.S.A.* **99** (16): 10237–10239. doi:10.1073/pnas.172399499. PMC 124895. PMID 12149485.
- [76] Mehagnoul-Schipper, DJ; van der Kallen, BF; Colier, WNJM; van der Sluijs, MC; van Erning, LJ; Thijssen, HO; Oeseburg, B; Hoefnagels, WH et al. (2002). "Simultaneous measurements of cerebral oxygenation changes during brain activation by near-infrared spectroscopy and functional magnetic resonance imaging in healthy young and elderly subjects.". *Hum Brain Mapp* **16** (1): 14–23. doi:10.1002/hbm.10026.
- [77] Soengas, JL; Aldegunde, M (2002). "Energy metabolism of fish brain". *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **131** (3): 271–296. doi:10.1016/S1096-4959(02)00022-2. PMID 11959012.
- [78] Carew, TJ (2000). "Ch. 1" (<http://books.google.com/?id=wEMTGwAACAAJ>). *Behavioral Neurobiology: the Cellular Organization of Natural Behavior*. Sinauer Associates. ISBN 978-0-87893-092-0. .
- [79] Churchland, PS; Koch, C; Sejnowski, TJ (1993). "What is computational neuroscience?". In Schwartz EL. *Computational Neuroscience*. MIT Press. pp. 46–55. ISBN 978-0-262-69164-2.



- [80] von Neumann, J; Churchland, PM; Churchland, PS (2000). *The Computer and the Brain*. Yale University Press. pp. xi–xxii. ISBN 978-0-300-08473-3.
- [81] Lettvin, JY; Maturana, HR; McCulloch, WS; Pitts, WH (1959). "What the frog's eye tells the frog's brain" (<http://jerome.lettvin.info/lettvin/Jerome/WhatTheFrogsEyeTellsTheFrogsBrain.pdf>) (pdf). *Proceedings of the Institute of Radio Engineering* **47**: 1940–1951. .
- [82] Hubel, DH; Wiesel, TN (2005). *Brain and visual perception: the story of a 25-year collaboration*. Oxford University Press US. pp. 657–704. ISBN 978-0-19-517618-6.
- [83] Farah, MJ (2000). *The Cognitive Neuroscience of Vision*. Wiley-Blackwell. pp. 1–29. ISBN 978-0-631-21403-8.
- [84] Engel, AK; Singer, W (2001). "Temporal binding and the neural correlates of sensory awareness". *Trends in Cognitive Sciences* **5** (1): 16–25. doi:10.1016/S1364-6613(00)01568-0. PMID 11164732.
- [85] Dayan, P; Abbott, LF (2005). "Ch.7: Network models". *Theoretical Neuroscience*. MIT Press. ISBN 978-0-262-54185-5.
- [86] Averbeck, BB; Lee, D (2004). "Coding and transmission of information by neural ensembles". *Trends in Neurosciences* **27** (4): 225–230. doi:10.1016/j.tins.2004.02.006. PMID 15046882.
- [87] *Principles of Neural Science*, Ch. 21
- [88] *Principles of Neural Science*, Ch. 34
- [89] *Principles of Neural Science*, Chs. 36, 37
- [90] *Principles of Neural Science*, Ch. 33
- [91] Dafny, N. "Anatomy of the spinal cord" (<http://neuroscience.uth.tmc.edu/s2/chapter03.html>). Neuroscience Online. . Retrieved 2011-10-10.
- [92] Dragoi, V. "Ocular motor system" (<http://neuroscience.uth.tmc.edu/s3/chapter07.html>). Neuroscience Online. . Retrieved 2011-10-10.
- [93] Gurney, K; Prescott, TJ; Wickens, JR; Redgrave, P (2004). "Computational models of the basal ganglia: from robots to membranes". *Trends in Neurosciences* **27** (8): 453–459. doi:10.1016/j.tins.2004.06.003. PMID 15271492.
- [94] *Principles of Neural Science*, Ch. 38
- [95] Shima, K; Tanji, J (1998). "Both supplementary and presupplementary motor areas are crucial for the temporal organization of multiple movements" (<http://jn.physiology.org/cgi/content/abstract/80/6/3247>). *Journal of Neurophysiology* **80** (6): 3247–3260. PMID 9862919.
- [96] Miller, EK; Cohen, JD (2001). "An integrative theory of prefrontal cortex function". *Annual Review of Neuroscience* **24** (1): 167–202. doi:10.1146/annurev.neuro.24.1.167. PMID 11283309.
- [97] *Principles of Neural Science*, Ch. 49
- [98] *Principles of Neural Science*, Ch. 45
- [99] Antle, MC; Silver, R (2005). "Orchestrating time: arrangements of the brain circadian clock" ([http://www.columbia.edu/cu/psychology/silver/publications2/149\\_antle\\_et\\_al.pdf](http://www.columbia.edu/cu/psychology/silver/publications2/149_antle_et_al.pdf)) (PDF). *Trends in Neurosciences* **28** (3): 145–151. doi:10.1016/j.tins.2005.01.003. PMID 15749168. .
- [100] *Principles of Neural Science*, Ch. 47
- [101] Kleitman, N (1938, revised 1963, reprinted 1987). *Sleep and Wakefulness*. The University of Chicago Press, Midway Reprints series. ISBN 0-226-44073-7.
- [102] Dougherty, P. "Hypothalamus: structural organization" (<http://neuroscience.uth.tmc.edu/s4/chapter01.html>). Neuroscience Online. . Retrieved 2011-10-11.
- [103] Gross, CG (1998). "Claude Bernard and the constancy of the internal environment" ([http://www.princeton.edu/~cggross/Neuroscientist\\_98\\_Bernard.pdf](http://www.princeton.edu/~cggross/Neuroscientist_98_Bernard.pdf)) (PDF). *The Neuroscientist* **4** (5): 380–385. doi:10.1177/107385849800400520. .
- [104] Dougherty, P. "Hypothalamic control of pituitary hormone" (<http://neuroscience.uth.tmc.edu/s4/chapter02.html>). Neuroscience Online. . Retrieved 2011-10-11.
- [105] Chiel, HJ; Beer, RD (1997). "The brain has a body: adaptive behavior emerges from interactions of nervous system, body, and environment". *Trends in Neurosciences* **20** (12): 553–557. doi:10.1016/S0166-2236(97)01149-1. PMID 9416664.
- [106] Berridge, KC (2004). "Motivation concepts in behavioral neuroscience". *Physiology & Behavior* **8** (2): 179–209. doi:10.1016/j.physbeh.2004.02.004. PMID 15159167.
- [107] Ardiel, EL; Rankin, CH (2010). "An elegant mind: learning and memory in *Caenorhabditis elegans*" (<http://learnmem.cshlp.org/content/17/4/191.long>). *Learning and Memory* **17** (4): 191–201. doi:10.1101/lm.960510. PMID 20335372. .
- [108] Hyman, SE; Malenka, RC (2001). "Addiction and the brain: the neurobiology of compulsion and its persistence". *Nature Reviews Neuroscience* **2** (10): 695–703. doi:10.1038/35094560. PMID 11584307.
- [109] Ramón y Cajal, S (1894). "The Croonian Lecture: La Fine Structure des Centres Nerveux". *Proceedings of the Royal Society of London* **55** (331–335): 444–468. doi:10.1098/rspl.1894.0063.
- [110] Lømo, T (2003). "The discovery of long-term potentiation". *Phil. Trans. Roy. Soc. London B (Biological Sciences)* **358** (1432): 617–620. doi:10.1098/rstb.2002.1226. PMC 1693150. PMID 12740104.
- [111] Malenka, R; Bear, M (2004). "LTP and LTD: an embarrassment of riches". *Neuron* **44** (1): 5–21. doi:10.1016/j.neuron.2004.09.012. PMID 15450156.
- [112] Curtis, CE; D'Esposito, M (2003). "Persistent activity in the prefrontal cortex during working memory". *Trends in Cognitive Sciences* **7** (9): 415–423. doi:10.1016/S1364-6613(03)00197-9. PMID 12963473.
- [113] Tulving, E; Markowitsch, HJ (1998). "Episodic and declarative memory: role of the hippocampus". *Hippocampus* **8** (3): 198–204. doi:10.1002/(SICI)1098-1063(1998)8:3<198::AID-HIPO2>3.0.CO;2-G. PMID 9662134.

- [114] Martin, A; Chao, LL (2001). "Semantic memory and the brain: structures and processes". *Current Opinion in Neurobiology* **11** (2): 194–201. doi:10.1016/S0959-4388(00)00196-3. PMID 11301239.
- [115] Balleine, BW; Liljeholm, Mimi; Ostlund, SB (2009). "The integrative function of the basal ganglia in instrumental learning". *Behavioral Brain Research* **199** (1): 43–52. doi:10.1016/j.bbr.2008.10.034. PMID 19027797.
- [116] Doya, K (2000). "Complementary roles of basal ganglia and cerebellum in learning and motor control". *Current Opinion in Neurobiology* **10** (6): 732–739. doi:10.1016/S0959-4388(00)00153-7. PMID 11240282.
- [117] *Principles of Neural Development*, Ch. 1
- [118] *Principles of Neural Development*, Ch. 4
- [119] *Principles of Neural Development*, Chs. 5, 7
- [120] *Principles of Neural Development*, Ch. 12
- [121] Wong, R (1999). "Retinal waves and visual system development". *Annual Review of Neuroscience* **22**: 29–47. doi:10.1146/annurev.neuro.22.1.29. PMID 10202531.
- [122] *Principles of Neural Development*, Ch. 6
- [123] Rakic, P (2002). "Adult neurogenesis in mammals: an identity crisis". *J. Neuroscience* **22** (3): 614–618. PMID 11826088.
- [124] Ridley, M (2003). *Nature via Nurture: Genes, Experience, and What Makes Us Human* (<http://books.google.com/?id=9TkUHQAACA AJ>). Forth Estate. pp. 1–6. ISBN 978-0-06-000678-5. .
- [125] Wiesel, T (1982). "Postnatal development of the visual cortex and the influence of environment" (<http://www.nobel.se/medicine/laureates/1981/wiesel-lecture.pdf>) (PDF). *Nature* **299** (5884): 583–591. doi:10.1038/299583a0. PMID 6811951. .
- [126] van Praag, H; Kempermann, G; Gage, FH (2000). "Neural consequences of environmental enrichment". *Nature Reviews Neuroscience* **1** (3): 191–198. doi:10.1038/35044558. PMID 11257907.
- [127] *Principles of Neural Science*, Ch. 1
- [128] Storrow, HA (1969). *Outline of Clinical Psychiatry* (<http://books.google.com/books?ei=pQ-ZTtmvG7TOiAK09rgn>). Appleton-Century-Crofts. pp. 27–30. .
- [129] Thagard, P (2008). "Cognitive Science" (<http://plato.stanford.edu/archives/fall2008/entries/cognitive-science/>). In Zalta, EN. The Stanford Encyclopedia of Philosophy. . Retrieved 2011-10-14.
- [130] Bear, MF; Connors, BW; Paradiso, MA (2007). "Ch. 2". *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins. ISBN 978-0-7817-6003-4.
- [131] Dowling, JE (2001). *Neurons and Networks*. Harvard University Press. pp. 15–24. ISBN 978-0-674-00462-7.
- [132] Wyllie, E; Gupta, A; Lachhwani, DK (2005). "Ch. 77". *The Treatment of Epilepsy: Principles and Practice*. Lippincott Williams & Wilkins. ISBN 978-0-7817-4995-4.
- [133] Laureys, S; Boly, M; Tononi, G (2009). "Functional neuroimaging". In Laureys S, Tononi G. *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*. Academic Press. pp. 31–42. ISBN 978-0-12-374168-4.
- [134] Carmena, JM et al. (2003). "Learning to Control a Brain–Machine Interface for Reaching and Grasping by Primates". *PLoS Biology* **1** (2): 193–208. doi:10.1371/journal.pbio.0000042. PMC 261882. PMID 14624244.
- [135] Kolb, B; Whishaw, I (2008). "Ch. 1". *Fundamentals of Human Neuropsychology*. Macmillan. ISBN 978-0-7167-9586-5.
- [136] Abbott, LF; Dayan, P (2001). "Preface". *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press. ISBN 978-0-262-54185-5.
- [137] Tonegawa, S; Nakazawa, K; Wilson, MA (2003). "Genetic neuroscience of mammalian learning and memory". *Phil. Trans. Roy. Soc. London B (Biological Sciences)* **358** (1432): 787–795. doi:10.1098/rstb.2002.1243. PMC 1693163. PMID 12740125.
- [138] Finger, S (2001). *Origins of Neuroscience*. Oxford University Press. pp. 14–15. ISBN 978-0-19-514694-3.
- [139] Finger, S (2001). *Origins of Neuroscience*. Oxford University Press. pp. 193–195. ISBN 978-0-19-514694-3.
- [140] Bloom, FE (1975). Schmidt FO, Worden FG, Swazey JP, Adelman G. ed. *The Neurosciences, Paths of Discovery*. MIT Press. p. 211. ISBN 978-0-262-23072-8.
- [141] Shepherd, GM (1991). "Ch.1 : Introduction and Overview". *Foundations of the Neuron Doctrine*. Oxford University Press. ISBN 978-0-19-506491-9.
- [142] Piccolino, M (2002). "Fifty years of the Hodgkin-Huxley era". *Trends in Neurosciences* **25** (11): 552–553. doi:10.1016/S0166-2236(02)02276-2. PMID 12392928.
- [143] Sherrington, CS (1942). *Man on his nature*. Cambridge University Press. p. 178. ISBN 978-0-8385-7701-1.
- [144] Jones, EG; Mendell, LM (1999). "Assessing the Decade of the Brain" (<http://www.sciencemag.org/cgi/content/summary/284/5415/739>). *Science* **284** (5415): 739. doi:10.1126/science.284.5415.739. PMID 10336393. .
- [145] Buzsáki, G (2004). "Large-scale recording of neuronal ensembles" (<http://osiris.rutgers.edu/BuzsakiHP/Publications/PDFs/Buzsaki2004NatNeurosci.pdf>). *Nature Neuroscience* **7** (5): 446–451. doi:10.1038/nn1233. PMID 15114356. .
- [146] Geschwind, DH; Konopka, G (2009). "Neuroscience in the era of functional genomics and systems biology". *Nature* **461** (7266): 908–915. doi:10.1038/nature08537. PMID 19829370.

## Further reading

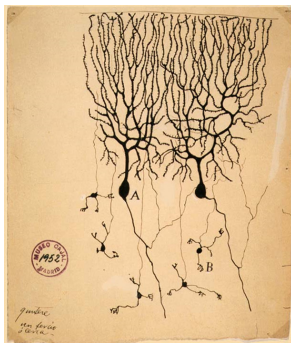
- Kandel, ER; Schwartz, JH; Jessel, TM (2000). *Principles of Neural Science*. McGraw-Hill Professional. ISBN 978-0-8385-7701-1.
- Purves, D; Lichtman, J (1985). *Principles of Neural Development* (<http://books.google.com/?id=t9JqAAAAMAAJ>). Sinauer Associates. ISBN 978-0-87893-744-8.

## External links

- Brain Museum (<http://brainmuseum.org/>), comparative mammalian brain collection
- BrainInfo (<http://braininfo.rprc.washington.edu>), neuroanatomy database
- Neuroscience for Kids (<http://faculty.washington.edu/chudler/neurok.html>)
- BrainMaps.org (<http://www.brainmaps.org/>), interactive high-resolution digital brain atlas of primate and non-primate brains
- The Brain from Top to Bottom (<http://thebrain.mcgill.ca>), at McGill University
- The HOPES Brain Tutorial (<http://www.stanford.edu/group/hopes/cgi-bin/wordpress/?p=3787>), at Stanford University

# Neuron

### *Neuron: nerve cell*



Drawing by Santiago Ramón y Cajal of neurons in the pigeon cerebellum. (A) Denotes Purkinje cells, an example of a multipolar neuron. (B) Denotes granule cells, which are also multipolar.

<b>NeuroLex ID</b>	<i>sao1417703748</i> <sup>[1]</sup>
--------------------	-------------------------------------

A **neuron** (pron.: /ˈnjʊərɒn/ *NYEWR-on* or pron.: /ˈnʊərɒn/ *NEWR-on*; also known as a **neurone** or **nerve cell**) is an electrically excitable cell that processes and transmits information through electrical and chemical signals. A chemical signal occurs via a synapse, a specialized connection with other cells. Neurons connect to each other to form neural networks. Neurons are the core components of the nervous system, which includes the brain, spinal cord, and peripheral ganglia. A number of specialized types of neurons exist: sensory neurons respond to touch, sound, light and numerous other stimuli affecting cells of the sensory organs that then send signals to the spinal cord and brain. Motor neurons receive signals from the brain and spinal cord, cause muscle contractions, and affect glands. Interneurons connect neurons to other neurons within the same region of the brain or spinal cord.

A typical neuron possesses a cell body (often called the soma), dendrites, and an axon. Dendrites are thin structures that arise from the cell body, often extending for hundreds of micrometres and branching multiple times, giving rise to a complex "dendritic tree". An axon is a special cellular extension that arises from the cell body at a site called the axon hillock and travels for a distance, as far as 1 meter in humans or even more in other species. The cell body of a neuron frequently gives rise to multiple dendrites, but never to more than one axon, although the axon may branch

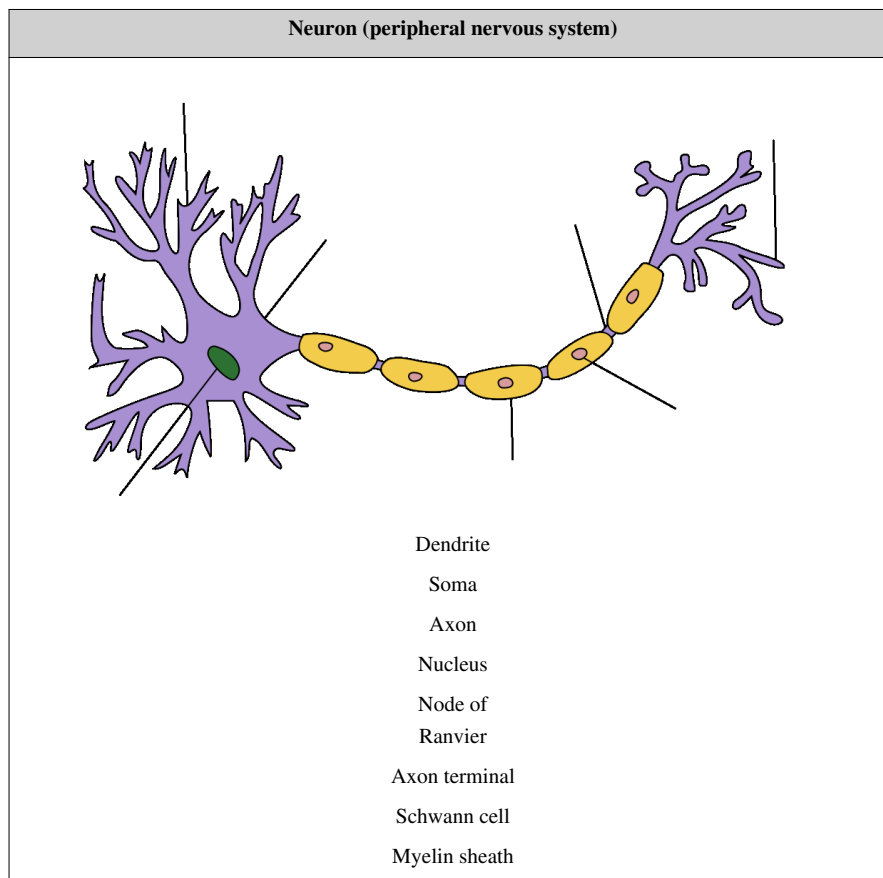
hundreds of times before it terminates. At the majority of synapses, signals are sent from the axon of one neuron to a dendrite of another. There are, however, many exceptions to these rules: neurons that lack dendrites, neurons that have no axon, synapses that connect an axon to another axon or a dendrite to another dendrite, etc.

All neurons are electrically excitable, maintaining voltage gradients across their membranes by means of metabolically driven ion pumps, which combine with ion channels embedded in the membrane to generate intracellular-versus-extracellular concentration differences of ions such as sodium, potassium, chloride, and calcium. Changes in the cross-membrane voltage can alter the function of voltage-dependent ion channels. If the voltage changes by a large enough amount, an all-or-none electrochemical pulse called an action potential is generated, which travels rapidly along the cell's axon, and activates synaptic connections with other cells when it arrives.

Neurons do not undergo cell division. In most cases, neurons are generated by special types of stem cells. Astrocytes, a type of glial cell, have also been observed to turn into neurons by virtue of the stem cell characteristic pluripotency. In humans, neurogenesis largely ceases during adulthood—but in two brain areas, the hippocampus and olfactory bulb, there is strong evidence for generation of substantial numbers of new neurons.<sup>[2][3]</sup>

## Overview

### Structure of a typical neuron



A neuron is a specialized type of cell found in the bodies of most animals (all members of the group Eumetazoa). Only sponges and a few other simpler animals have no neurons. The features that define a neuron are electrical excitability and the presence of synapses, which are complex membrane junctions that transmit signals to other cells. The body's neurons, plus the glial cells that give them structural and metabolic support, together constitute the nervous system. In vertebrates, the majority of neurons belong to the central nervous system, but some reside in peripheral ganglia, and many sensory neurons are situated in sensory organs such as the retina and cochlea.

Although neurons are very diverse and there are exceptions to nearly every rule, it is convenient to begin with a schematic description of the structure and function of a "typical" neuron. A typical neuron is divided into three parts: the soma or cell body, dendrites, and axon. The soma is usually compact; the axon and dendrites are filaments that extrude from it. Dendrites typically branch profusely, getting thinner with each branching, and extending their farthest branches a few hundred micrometres from the soma. The axon leaves the soma at a swelling called the axon hillock, and can extend for great distances, giving rise to hundreds of branches. Unlike dendrites, an axon usually maintains the same diameter as it extends. The soma may give rise to numerous dendrites, but never to more than one axon. Synaptic signals from other neurons are received by the soma and dendrites; signals to other neurons are transmitted by the axon. A typical synapse, then, is a contact between the axon of one neuron and a dendrite or soma of another. Synaptic signals may be excitatory or inhibitory. If the net excitation received by a neuron over a short period of time is large enough, the neuron generates a brief pulse called an action potential, which originates at the soma and propagates rapidly along the axon, activating synapses onto other neurons as it goes.

Many neurons fit the foregoing schema in every respect, but there are also exceptions to most parts of it. There are no neurons that lack a soma, but there are neurons that lack dendrites, and others that lack an axon. Furthermore, in addition to the typical axodendritic and axosomatic synapses, there are axoaxonic (axon-to-axon) and dendrodendritic (dendrite-to-dendrite) synapses.

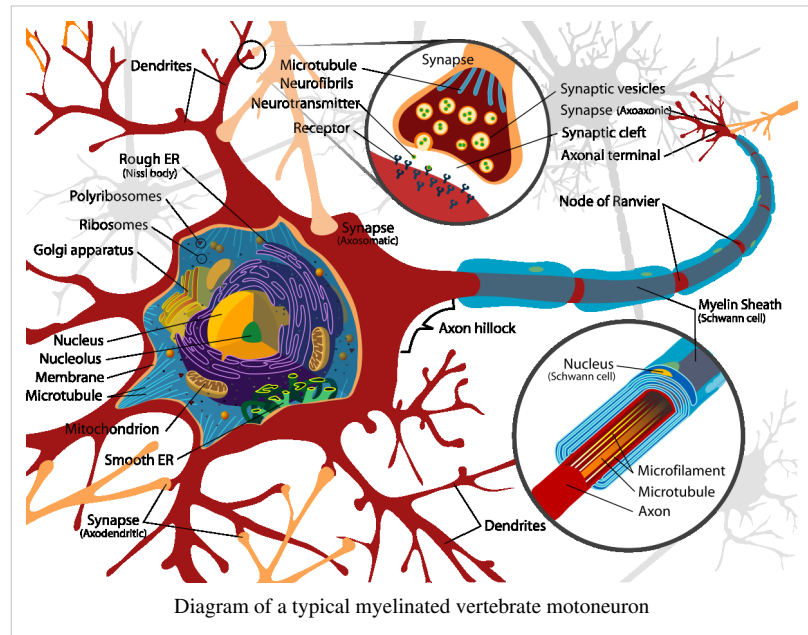
The key to neural function is the synaptic signaling process, which is partly electrical and partly chemical. The electrical aspect depends on properties of the neuron's membrane. Like all animal cells, every neuron is surrounded by a plasma membrane, a bilayer of lipid molecules with many types of protein structures embedded in it. A lipid bilayer is a powerful electrical insulator, but in neurons, many of the protein structures embedded in the membrane are electrically active. These include ion channels that permit electrically charged ions to flow across the membrane, and ion pumps that actively transport ions from one side of the membrane to the other. Most ion channels are permeable only to specific types of ions. Some ion channels are voltage gated, meaning that they can be switched between open and closed states by altering the voltage difference across the membrane. Others are chemically gated, meaning that they can be switched between open and closed states by interactions with chemicals that diffuse through the extracellular fluid. The interactions between ion channels and ion pumps produce a voltage difference across the membrane, typically a bit less than 1/10 of a volt at baseline. This voltage has two functions: first, it provides a power source for an assortment of voltage-dependent protein machinery that is embedded in the membrane; second, it provides a basis for electrical signal transmission between different parts of the membrane.

Neurons communicate by chemical and electrical synapses in a process known as synaptic transmission. The fundamental process that triggers synaptic transmission is the action potential, a propagating electrical signal that is generated by exploiting the electrically excitable membrane of the neuron. This is also known as a wave of depolarization.

## Anatomy and histology

Neurons are highly specialized for the processing and transmission of cellular signals. Given the diversity of functions performed by neurons in different parts of the nervous system, there is, as expected, a wide variety in the shape, size, and electrochemical properties of neurons. For instance, the soma of a neuron can vary from 4 to 100 micrometers in diameter.<sup>[4]</sup>

- The soma is the central part of the neuron. It contains the nucleus of the cell, and therefore is where most protein synthesis occurs. The nucleus ranges from 3 to 18 micrometers in diameter.<sup>[5]</sup>



- The dendrites of a neuron are cellular extensions with many branches, and metaphorically this overall shape and structure is referred to as a dendritic tree. This is where the majority of input to the neuron occurs via the dendritic spine.
- The axon is a finer, cable-like projection that can extend tens, hundreds, or even tens of thousands of times the diameter of the soma in length. The axon carries nerve signals away from the soma (and also carries some types of information back to it). Many neurons have only one axon, but this axon may—and usually will—undergo extensive branching, enabling communication with many target cells. The part of the axon where it emerges from the soma is called the axon hillock. Besides being an anatomical structure, the axon hillock is also the part of the neuron that has the greatest density of voltage-dependent sodium channels. This makes it the most easily-excited part of the neuron and the spike initiation zone for the axon: in electrophysiological terms it has the most negative action potential threshold. While the axon and axon hillock are generally involved in information outflow, this region can also receive input from other neurons.
- The axon terminal contains synapses, specialized structures where neurotransmitter chemicals are released to communicate with target neurons.

Although the canonical view of the neuron attributes dedicated functions to its various anatomical components, dendrites and axons often act in ways contrary to their so-called main function.

Axons and dendrites in the central nervous system are typically only about one micrometer thick, while some in the peripheral nervous system are much thicker. The soma is usually about 10–25 micrometers in diameter and often is not much larger than the cell nucleus it contains. The longest axon of a human motoneuron can be over a meter long, reaching from the base of the spine to the toes. Sensory neurons have axons that run from the toes to the dorsal columns, over 1.5 meters in adults. Giraffes have single axons several meters in length running along the entire length of their necks. Much of what is known about axonal function comes from studying the squid giant axon, an ideal experimental preparation because of its relatively immense size (0.5–1 millimeters thick, several centimeters long).

Fully differentiated neurons are permanently postmitotic,<sup>[6]</sup> however, recent research shows that additional neurons throughout the brain can originate from neural stem cells found throughout the brain but in particularly high concentrations in the subventricular zone and subgranular zone through the process of neurogenesis.<sup>[7]</sup>



## Histology and internal structure

Nerve cell bodies stained with basophilic dyes show numerous microscopic clumps of Nissl substance (named after German psychiatrist and neuropathologist Franz Nissl, 1860–1919), which consists of rough endoplasmic reticulum and associated ribosomal RNA. The prominence of the Nissl substance can be explained by the fact that nerve cells are metabolically very active, and hence are involved in large amounts of protein synthesis.

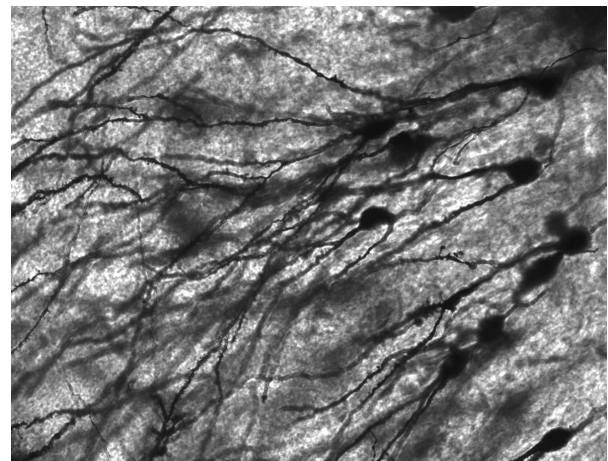
The cell body of a neuron is supported by a complex meshwork of structural proteins called neurofilaments, which are assembled into larger neurofibrils. Some neurons also contain pigment granules, such as neuromelanin (a brownish-black pigment, byproduct of synthesis of catecholamines) and lipofuscin (yellowish-brown pigment that accumulates with age).

There are different internal structural characteristics between axons and dendrites. Typical axons almost never contain ribosomes, except some in the initial segment. Dendrites contain granular endoplasmic reticulum or ribosomes, with diminishing amounts with distance from the cell body.

## Classes

Neurons exist in a number of different shapes and sizes and can be classified by their morphology and function. The anatomist Camillo Golgi grouped neurons into two types; type I with long axons used to move signals over long distances and type II with short axons, which can often be confused with dendrites. Type I cells can be further divided by where the cell body or soma is located. The basic morphology of type I neurons, represented by spinal motor neurons, consists of a cell body called the soma and a long thin axon covered by the myelin sheath. Around the cell body is a branching dendritic tree that receives signals from other neurons. The end of the axon has branching terminals (axon terminal) that release neurotransmitters into a gap called the synaptic cleft between the terminals and the dendrites of the next neuron.

## Structural classification



Golgi-stained neurons in human hippocampal tissue

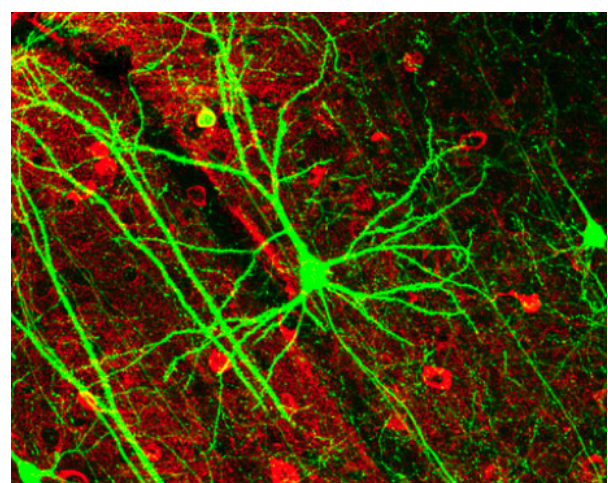
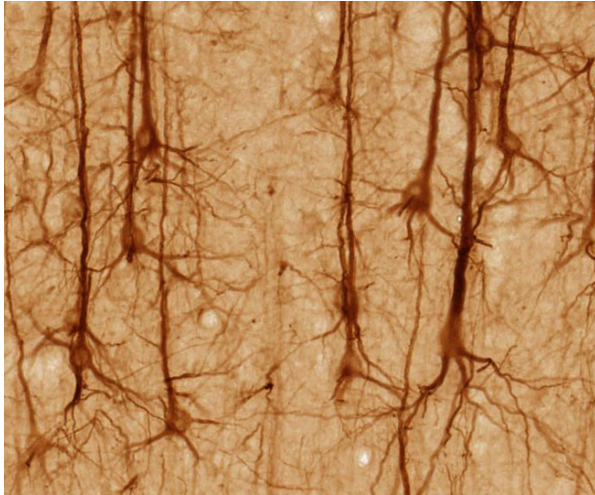


Image of pyramidal neurons in mouse cerebral cortex expressing green fluorescent protein. The red staining indicates GABAergic interneurons.<sup>[8]</sup>





SMI32-stained pyramidal neurons in cerebral cortex

### Polarity

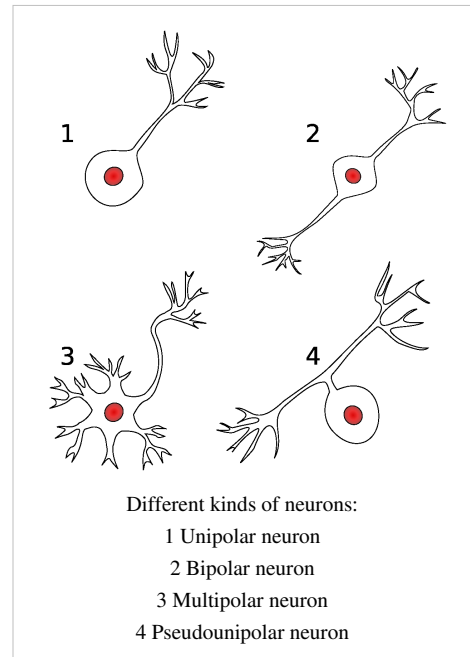
Most neurons can be anatomically characterized as:

- Unipolar or pseudounipolar: dendrite and axon emerging from same process.
- Bipolar: axon and single dendrite on opposite ends of the soma.
- Multipolar: more than two dendrites:
  - Golgi I: neurons with long-projecting axonal processes; examples are pyramidal cells, Purkinje cells, and anterior horn cells.
  - Golgi II: neurons whose axonal process projects locally; the best example is the granule cell.

### Other

Furthermore, some unique neuronal types can be identified according to their location in the nervous system and distinct shape. Some examples are:

- Basket cells, interneurons that form a dense plexus of terminals around the soma of target cells, found in the cortex and cerebellum.
- Betz cells, large motor neurons.
- Medium spiny neurons, most neurons in the corpus striatum.
- Purkinje cells, huge neurons in the cerebellum, a type of Golgi I multipolar neuron.
- Pyramidal cells, neurons with triangular soma, a type of Golgi I.
- Renshaw cells, neurons with both ends linked to alpha motor neurons.
- Granule cells, a type of Golgi II neuron.
- Anterior horn cells, motoneurons located in the spinal cord.
- Spindle cells, interneurons that connect widely separated areas of the brain



## Functional classification

### Direction

- Afferent neurons convey information from tissues and organs into the central nervous system and are sometimes also called sensory neurons.
- Efferent neurons transmit signals from the central nervous system to the effector cells and are sometimes called motor neurons.
- Interneurons connect neurons within specific regions of the central nervous system.

Afferent and efferent also refer generally to neurons that, respectively, bring information to or send information from the brain region.

### Action on other neurons

A neuron affects other neurons by releasing a neurotransmitter that binds to chemical receptors. The effect upon the postsynaptic neuron is determined not by the presynaptic neuron or by the neurotransmitter, but by the type of receptor that is activated. A neurotransmitter can be thought of as a key, and a receptor as a lock: the same type of key can here be used to open many different types of locks. Receptors can be classified broadly as *excitatory* (causing an increase in firing rate), *inhibitory* (causing a decrease in firing rate), or *modulatory* (causing long-lasting effects not directly related to firing rate).

The two most common neurotransmitters in the brain, glutamate and GABA, have actions that are largely consistent. Glutamate acts on several different types of receptors, and have effects that are excitatory at ionotropic receptors and a modulatory effect at metabotropic receptors. Similarly GABA acts on several different types of receptors, but all of them have effects (in adult animals, at least) that are inhibitory. Because of this consistency, it is common for neuroscientists to simplify the terminology by referring to cells that release glutamate as "excitatory neurons," and cells that release GABA as "inhibitory neurons." Since over 90% of the neurons in the brain release either glutamate or GABA, these labels encompass the great majority of neurons. There are also other types of neurons that have consistent effects on their targets, for example "excitatory" motor neurons in the spinal cord that release acetylcholine, and "inhibitory" spinal neurons that release glycine.

The distinction between excitatory and inhibitory neurotransmitters is not absolute, however. Rather, it depends on the class of chemical receptors present on the postsynaptic neuron. In principle, a single neuron, releasing a single neurotransmitter, can have excitatory effects on some targets, inhibitory effects on others, and modulatory effects on others still. For example, photoreceptor cells in the retina constantly release the neurotransmitter glutamate in the absence of light. So-called OFF bipolar cells are, like most neurons, excited by the released glutamate. However, neighboring target neurons called ON bipolar cells are instead *inhibited* by glutamate, because they lack the typical ionotropic glutamate receptors and instead express a class of inhibitory metabotropic glutamate receptors.<sup>[9]</sup> When light is present, the photoreceptors cease releasing glutamate, which relieves the ON bipolar cells from inhibition, activating them; this simultaneously removes the excitation from the OFF bipolar cells, silencing them.

It is possible to identify the type of inhibitory effect a presynaptic neuron will have on a postsynaptic neuron, based on the proteins the presynaptic neuron expresses. Parvalbumin-expressing neurons typically dampen the output signal of the postsynaptic neuron in the visual cortex, whereas somatostatin-expressing neurons typically block dendritic inputs to the postsynaptic neuron.<sup>[10]</sup>

### Discharge patterns

Neurons can be classified according to their electrophysiological characteristics:

- **Tonic or regular spiking.** Some neurons are typically constantly (or tonically) active. Example: interneurons in neurostriatum.
- **Phasic or bursting.** Neurons that fire in bursts are called phasic.
- **Fast spiking.** Some neurons are notable for their high firing rates, for example some types of cortical inhibitory interneurons, cells in globus pallidus, retinal ganglion cells.<sup>[11][12]</sup>

### Classification by neurotransmitter production

Neurons differ in the type of neurotransmitter they manufacture. Some examples are:

- **Cholinergic neurons—acetylcholine.** Acetylcholine is released from presynaptic neurons into the synaptic cleft. It acts as a ligand for both ligand-gated ion channels and metabotropic (GPCRs) muscarinic receptors. Nicotinic receptors, are pentameric ligand-gated ion channels composed of alpha and beta subunits that bind nicotine. Ligand binding opens the channel causing influx of  $\text{Na}^+$  depolarization and increases the probability of presynaptic neurotransmitter release.
- **GABAergic neurons—gamma aminobutyric acid.** GABA is one of two neuroinhibitors in the CNS, the other being Glycine. GABA has a homologous function to ACh, gating anion channels that allow  $\text{Cl}^-$  ions to enter the post synaptic neuron.  $\text{Cl}^-$  causes hyperpolarization within the neuron, decreasing the probability of an action potential firing as the voltage becomes more negative (recall that for an action potential to fire, a positive voltage threshold must be reached).
- **Glutamatergic neurons—glutamate.** Glutamate is one of two primary excitatory amino acids, the other being Aspartate. Glutamate receptors are one of four categories, three of which are ligand-gated ion channels and one of which is a G-protein coupled receptor (often referred to as GPCR).
  1. AMPA and Kainate receptors both function as cation channels permeable to  $\text{Na}^+$  cation channels mediating fast excitatory synaptic transmission
  2. NMDA receptors are another cation channel that is more permeable to  $\text{Ca}^{2+}$ . The function of NMDA receptors is dependant on Glycine receptor binding as a co-agonist within the channel pore. NMDA receptors do not function without both ligands present.
  3. Metabotropic receptors, GPCRs modulate synaptic transmission and postsynaptic excitability.
 

Glutamate can cause excitotoxicity when blood flow to the brain is interrupted, resulting in brain damage. When blood flow is suppressed, glutamate is released from presynaptic neurons causing NMDA and AMPA receptor activation moreso than would normally be the case outside of stress conditions, leading to elevated  $\text{Ca}^{2+}$  and  $\text{Na}^+$  entering the post synaptic neuron and cell damage.
- **Dopaminergic neurons—dopamine.** Dopamine is a neurotransmitter that acts on D1 type (D1 and D5) Gs coupled receptors, which increase cAMP and PKA, and D2 type (D2, D3, and D4) receptors, which activate Gi-coupled receptors that decrease cAMP and PKA. Dopamine is connected to mood and behavior, and modulates both pre and post synaptic neurotransmission. Loss of dopamine neurons in the substantia nigra has been linked to Parkinson's disease.
- **Serotonergic neurons—serotonin.** Serotonin,(5-Hydroxytryptamine, 5-HT), can act as excitatory or inhibitory. Of the four 5-HT receptor classes, 3 are GPCR and 1 is ligand gated cation channel. Serotonin is synthesized from tryptophan by tryptophan hydroxylase, and then further by aromatic acid decarboxylase. A lack of 5-HT at postsynaptic neurons has been linked to depression. Drugs that block the presynaptic serotonin transporter are used for treatment, such as Prozac and Zoloft.

## Connectivity

Neurons communicate with one another via synapses, where the axon terminal or *en passant* boutons (terminals located along the length of the axon) of one cell impinges upon another neuron's dendrite, soma or, less commonly, axon. Neurons such as Purkinje cells in the cerebellum can have over 1000 dendritic branches, making connections with tens of thousands of other cells; other neurons, such as the magnocellular neurons of the supraoptic nucleus, have only one or two dendrites, each of which receives thousands of synapses. Synapses can be excitatory or inhibitory and either increase or decrease activity in the target neuron. Some neurons also communicate via electrical synapses, which are direct, electrically-conductive junctions between cells.

In a chemical synapse, the process of synaptic transmission is as follows: when an action potential reaches the axon terminal, it opens voltage-gated calcium channels, allowing calcium ions to enter the terminal. Calcium causes synaptic vesicles filled with neurotransmitter molecules to fuse with the membrane, releasing their contents into the synaptic cleft. The neurotransmitters diffuse across the synaptic cleft and activate receptors on the postsynaptic neuron.

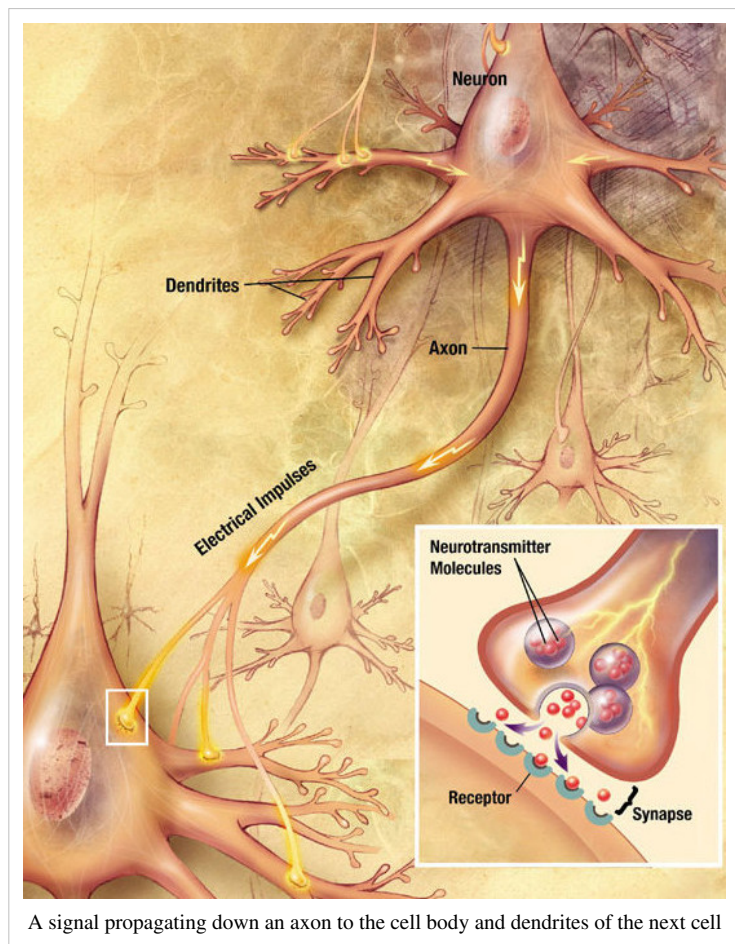
The human brain has a huge number of synapses. Each of the  $10^{11}$  (one hundred billion) neurons has on average 7,000 synaptic connections to other neurons. It has been estimated that the brain of a three-year-old child has about  $10^{15}$  synapses (1 quadrillion). This number declines with age, stabilizing by adulthood. Estimates vary for an adult, ranging from  $10^{14}$  to  $5 \times 10^{14}$  synapses (100 to 500 trillion).<sup>[13]</sup>

## Mechanisms for propagating action potentials

In 1937, John Zachary Young suggested that the squid giant axon could be used to study neuronal electrical properties.<sup>[14]</sup> Being larger than but similar in nature to human neurons, squid cells were easier to study. By inserting electrodes into the giant squid axons, accurate measurements were made of the membrane potential.

The cell membrane of the axon and soma contain voltage-gated ion channels that allow the neuron to generate and propagate an electrical signal (an action potential). These signals are generated and propagated by charge-carrying ions including sodium ( $\text{Na}^+$ ), potassium ( $\text{K}^+$ ), chloride ( $\text{Cl}^-$ ), and calcium ( $\text{Ca}^{2+}$ ).

There are several stimuli that can activate a neuron leading to electrical activity, including pressure, stretch, chemical transmitters, and changes of the electric potential across the cell membrane.<sup>[15]</sup> Stimuli cause specific ion-channels within the cell membrane to open, leading to a flow of ions through the cell membrane, changing the membrane potential.



A signal propagating down an axon to the cell body and dendrites of the next cell

Thin neurons and axons require less metabolic expense to produce and carry action potentials, but thicker axons convey impulses more rapidly. To minimize metabolic expense while maintaining rapid conduction, many neurons have insulating sheaths of myelin around their axons. The sheaths are formed by glial cells: oligodendrocytes in the central nervous system and Schwann cells in the peripheral nervous system. The sheath enables action potentials to travel faster than in unmyelinated axons of the same diameter, whilst using less energy. The myelin sheath in peripheral nerves normally runs along the axon in sections about 1 mm long, punctuated by unsheathed nodes of Ranvier, which contain a high density of voltage-gated ion channels. Multiple sclerosis is a neurological disorder that results from demyelination of axons in the central nervous system.

Some neurons do not generate action potentials, but instead generate a graded electrical signal, which in turn causes graded neurotransmitter release. Such nonspiking neurons tend to be sensory neurons or interneurons, because they cannot carry signals long distances.

## Neural coding

Neural coding is concerned with how sensory and other information is represented in the brain by neurons. The main goal of studying neural coding is to characterize the relationship between the stimulus and the individual or ensemble neuronal responses, and the relationships amongst the electrical activities of the neurons within the ensemble.<sup>[16]</sup> It is thought that neurons can encode both digital and analog information.<sup>[17]</sup>

## All-or-none principle

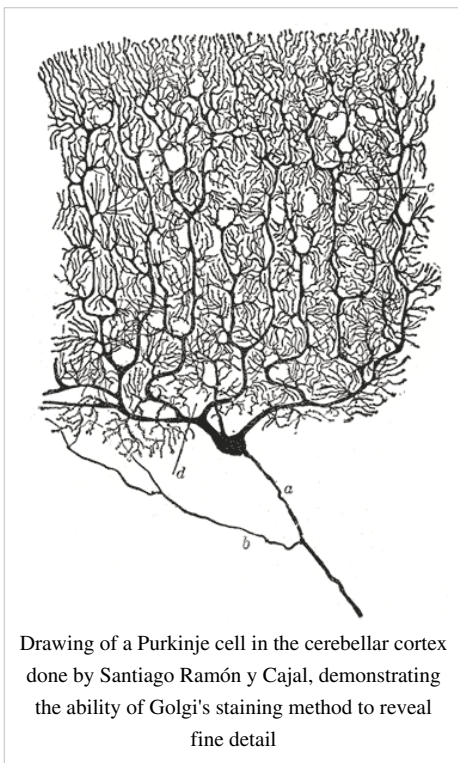
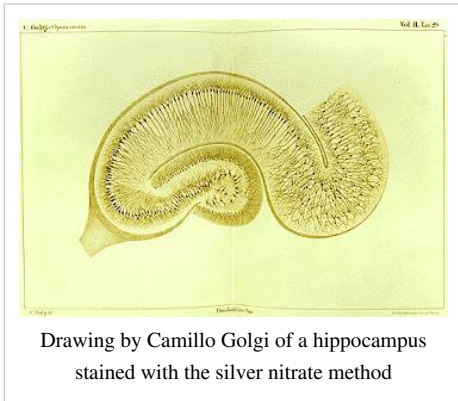
The conduction of nerve impulses is an example of an all-or-none response. In other words, if a neuron responds at all, then it must respond completely. Greater intensity of stimulation does not produce a stronger signal but can produce a higher frequency of firing. There are different types of receptor response to stimulus, slowly adapting or tonic receptors respond to steady stimulus and produce a steady rate of firing. These tonic receptors most often respond to increased intensity of stimulus by increasing their firing frequency, usually as a power function of stimulus plotted against impulses per second. This can be likened to an intrinsic property of light where to get greater intensity of a specific frequency (color) there have to be more photons, as the photons can't become "stronger" for a specific frequency.

There are a number of other receptor types that are called quickly-adapting or phasic receptors, where firing decreases or stops with steady stimulus; examples include: skin when touched by an object causes the neurons to fire, but if the object maintains even pressure against the skin, the neurons stop firing. The neurons of the skin and muscles that are responsive to pressure and vibration have filtering accessory structures that aid their function.

The pacinian corpuscle is one such structure. It has concentric layers like an onion, which form around the axon terminal. When pressure is applied and the corpuscle is deformed, mechanical stimulus is transferred to the axon, which fires. If the pressure is steady, there is no more stimulus; thus, typically these neurons respond with a transient depolarization during the initial deformation and again when the pressure is removed, which causes the corpuscle to change shape again. Other types of adaptation are important in extending the function of a number of other neurons.<sup>[18]</sup>



## History



The term *neuron* was coined by the German anatomist Heinrich Wilhelm Waldeyer. The neuron's place as the primary functional unit of the nervous system was first recognized in the early 20th century through the work of the Spanish anatomist Santiago Ramón y Cajal.<sup>[19]</sup> Ramón y Cajal proposed that neurons were discrete cells that communicated with each other via specialized junctions, or spaces, between cells.<sup>[19]</sup> This became known as the neuron doctrine, one of the central tenets of modern neuroscience.<sup>[19]</sup> To observe the structure of individual neurons, Ramón y Cajal improved a silver staining process known as Golgi's method, which had been developed by his rival, Camillo Golgi.<sup>[19]</sup> Cajal's improvement, which involved a technique he called "double impregnation", is still in use. The silver impregnation stains are an extremely useful method for neuroanatomical investigations because, for reasons unknown, it stains a very small percentage of cells in a tissue, so one is able to see the complete micro structure of individual neurons without much overlap from other cells in the densely packed brain.<sup>[20]</sup>

### The neuron doctrine

The neuron doctrine is the now fundamental idea that neurons are the basic structural and functional units of the nervous system. The theory was put forward by Santiago Ramón y Cajal in the late 19th century. It held that neurons are discrete cells (not connected in a meshwork), acting as metabolically distinct units.

Later discoveries yielded a few refinements to the simplest form of the doctrine. For example, glial cells, which are not considered neurons, play an essential role in information processing.<sup>[21]</sup> Also, electrical synapses are more common than previously thought,<sup>[22]</sup> meaning that there are direct, cytoplasmic connections between neurons. In fact, there are examples of neurons forming even tighter coupling: the squid giant axon arises from the fusion of multiple axons.<sup>[23]</sup>

Ramón y Cajal also postulated the Law of Dynamic Polarization, which states that a neuron receives signals at its dendrites and cell body and transmits them, as action potentials, along the axon in one direction: away from the cell body.<sup>[24]</sup> The Law of Dynamic Polarization has important exceptions; dendrites can serve as synaptic output sites of neurons<sup>[25]</sup> and axons can receive synaptic inputs.<sup>[26]</sup>

### Neurons in the brain

The number of neurons in the brain varies dramatically from species to species.<sup>[27]</sup> One estimate puts the human brain at about 100 billion ( $10^{11}$ ) neurons and 100 trillion ( $10^{14}$ ) synapses.<sup>[27]</sup> A lower 2012 estimate is 86 billion neurons, of which 16.3 billion are in the cerebral cortex, and 69 billion in the cerebellum.<sup>[28]</sup> By contrast, the nematode worm *Caenorhabditis elegans* has just 302 neurons, making it an ideal experimental subject as scientists have been able to map all of the organism's neurons. The fruit fly *Drosophila melanogaster*, a common subject in biological experiments, has around 100,000 neurons and exhibits many complex behaviors. Many properties of

neurons, from the type of neurotransmitters used to ion channel composition, are maintained across species, allowing scientists to study processes occurring in more complex organisms in much simpler experimental systems.

## Neurological disorders

**Charcot–Marie–Tooth disease** (CMT), also known as *hereditary motor and sensory neuropathy* (HMSN), *hereditary sensorimotor neuropathy* and *peroneal muscular atrophy*, is a heterogeneous inherited disorder of nerves (neuropathy) that is characterized by loss of muscle tissue and touch sensation, predominantly in the feet and legs but also in the hands and arms in the advanced stages of disease. Presently incurable, this disease is one of the most common inherited neurological disorders, with 37 in 100,000 affected.

**Alzheimer's disease** (AD), also known simply as *Alzheimer's*, is a neurodegenerative disease characterized by progressive cognitive deterioration together with declining activities of daily living and neuropsychiatric symptoms or behavioral changes. The most striking early symptom is loss of short-term memory (amnesia), which usually manifests as minor forgetfulness that becomes steadily more pronounced with illness progression, with relative preservation of older memories. As the disorder progresses, cognitive (intellectual) impairment extends to the domains of language (aphasia), skilled movements (apraxia), and recognition (agnosia), and functions such as decision-making and planning become impaired.

**Parkinson's disease** (PD), also known as *Parkinson disease*, is a degenerative disorder of the central nervous system that often impairs the sufferer's motor skills and speech. Parkinson's disease belongs to a group of conditions called movement disorders. It is characterized by muscle rigidity, tremor, a slowing of physical movement (bradykinesia), and in extreme cases, a loss of physical movement (akinesia). The primary symptoms are the results of decreased stimulation of the motor cortex by the basal ganglia, normally caused by the insufficient formation and action of dopamine, which is produced in the dopaminergic neurons of the brain. Secondary symptoms may include high level cognitive dysfunction and subtle language problems. PD is both chronic and progressive.

**Myasthenia gravis** is a neuromuscular disease leading to fluctuating muscle weakness and fatigability during simple activities. Weakness is typically caused by circulating antibodies that block acetylcholine receptors at the post-synaptic neuromuscular junction, inhibiting the stimulative effect of the neurotransmitter acetylcholine. Myasthenia is treated with immunosuppressants, cholinesterase inhibitors and, in selected cases, thymectomy.

## Demyelination

Demyelination is the act of demyelinating, or the loss of the myelin sheath insulating the nerves. When myelin degrades, conduction of signals along the nerve can be impaired or lost, and the nerve eventually withers. This leads to certain neurodegenerative disorders like multiple sclerosis and chronic inflammatory demyelinating polyneuropathy.

## Axonal degeneration

Although most injury responses include a calcium influx signaling to promote resealing of severed parts, axonal injuries initially lead to acute axonal degeneration (AAD), which is rapid separation of the proximal and distal ends within 30 minutes of injury. Degeneration follows with swelling of the axolemma, and eventually leads to bead like formation. Granular disintegration of the axonal cytoskeleton and inner organelles occurs after axolemma degradation. Early changes include accumulation of mitochondria in the paranodal regions at the site of injury. Endoplasmic reticulum degrades and mitochondria swell up and eventually disintegrate. The disintegration is dependent on Ubiquitin and Calpain proteases (caused by influx of calcium ion), suggesting that axonal degeneration is an active process. Thus the axon undergoes complete fragmentation. The process takes about roughly 24 hrs in the PNS, and longer in the CNS. The signaling pathways leading to axolemma degeneration are currently unknown.



## Nerve regeneration

It has been demonstrated that neurogenesis can sometimes occur in the adult vertebrate brain, a finding that led to controversy in 1999.<sup>[29]</sup> However, more recent studies of the age of human neurons suggest that this process occurs only for a minority of cells, and the overwhelming majority of neurons comprising the neocortex were formed before birth and persist without replacement.<sup>[3]</sup>

It is often possible for peripheral axons to regrow if they are severed. A report in *Nature* suggested that researchers had found a way to transform human skin cells into working nerve cells using a process called transdifferentiation in which "cells are forced to adopt new identities."<sup>[30]</sup>

## References

- [1] <http://www.neurolex.org/wiki/sao1417703748>
- [2] Wade, Nicholas (1999-10-15). "Brain may grown new cells daily" (<http://www.nytimes.com/1999/10/15/us/brain-may-grow-new-cells-daily.html>). *The New York Times*. .
- [3] Nowakowski, R. S. (2006). "Stable neuron numbers from cradle to grave". *Proceedings of the National Academy of Sciences* **103** (33): 12219. doi:10.1073/pnas.0605605103.
- [4] Davies, Melissa (2002-04-09). "The Neuron: size comparison" (<http://www.ualberta.ca/~neuro/OnlineIntro/NeuronExample.htm>). *Neuroscience: A journey through the brain*. . Retrieved 2009-06-20.
- [5] Chudler, Eric H.. "Brain Facts and Figures" (<http://faculty.washington.edu/chudler/facts.html>). *Neuroscience for Kids*. . Retrieved 2009-06-20.
- [6] Herrup K, Yang Y (Ma3371545712287 Content-Dispositin2124 2007). "Cell cycle regulation in the postmitotic neuron: oxymoron or new biology?". *Nat. Rev. Neurosci.* **8** (5): 368–78. doi:10.1038/nrn2124. PMID 17453017.
- [7] Alvarez-Buylla A, Garcia-Verdugo JM (February 1, 2002). "Neurogenesis in adult subventricular zone" (<http://www.jneurosci.org/cgi/pmidlookup?view=long&pmid=11826091>). *Journal of Neuroscience* **22** (3): 629–34. PMID 11826091. . Retrieved 2009-06-20.
- [8] Lee, Wei-Chung Allen; Huang, Hayden; Feng, Guoping; Sanes, Joshua R.; Brown, Emery N.; So, Peter T.; Nedivi, Elly (2006). "Dynamic Remodeling of Dendritic Arbors in GABAergic Interneurons of Adult Visual Cortex". *PLoS Biology* **4** (2): e29. doi:10.1371/journal.pbio.0040029. PMC 1318477. PMID 16366735.
- [9] Gerber U (2003). "Metabotropic glutamate receptors in vertebrate retina" (<http://www.springerlink.com/content/m748132506x00lm4/>). *Doc Ophthalmol* **106** (1): 83–87. doi:10.1023/A:1022477203420. PMID 12675489. .
- [10] Wilson, Nathan R.; Runyan, Caroline A.; Wang, Forea L.; Sur, Mriganka (2012). "Division and subtraction by distinct cortical inhibitory networks in vivo". *Nature* **488** (7411): 343–8. doi:10.1038/nature11347. PMID 22878717.
- [11] Kolodin, YO; Veselovskaia, NN; Veselovsky, NS; Fedulova, SA. "Ion conductances related to shaping the repetitive firing in rat retinal ganglion cells" (<http://www.blackwellpublishing.com/aphmeeting/abstract.asp?MeetingID=&id=61198>). *Acta Physiologica Congress*. . Retrieved 2009-06-20.
- [12] "Ionic conductances underlying excitability in tonically firing retinal ganglion cells of adult rat" (<http://ykolodin.50webs.com/>). Ykolodin.50webs.com. 2008-04-27. . Retrieved 2013-02-16.
- [13] Drachman D (2005). "Do we have brain to spare?". *Neurology* **64** (12): 2004–5. doi:10.1212/01.WNL.0000166914.38327.BB. PMID 15985565.
- [14] Chudler, Eric H.. "Milestones in Neuroscience Research" (<http://faculty.washington.edu/chudler/hist.html>). *Neuroscience for Kids*. . Retrieved 2009-06-20.
- [15] Patlak, Joe; Gibbons, Ray (2000-11-01). "Electrical Activity of Nerves" (<http://physioweb.med.uvm.edu/cardiaccep/EP/nervecells.htm>). *Aps in Nerve Cells*. . Retrieved 2009-06-20.
- [16] Brown EN, Kass RE, and Mitra PP (2004). "Multiple neural spike train data analysis: State-of-the-art and future challenges". *Nature Neuroscience* **7** (5): 456–61. doi:10.1038/nn1228. PMID 15114358.
- [17] Thorpe, SJ (1990) Spike arrival times: A highly efficient coding scheme for neural networks ([http://web.archive.org/web/20120215151304/http://pop.cerco.ups-tlse.fr/fr\\_vers/documents/thorpe\\_sj\\_90\\_91.pdf](http://web.archive.org/web/20120215151304/http://pop.cerco.ups-tlse.fr/fr_vers/documents/thorpe_sj_90_91.pdf)). In R. Eckmiller, G. Hartmann, & G. Hauske (Eds.) *Parallel processing in neural systems*, Elsevier, pp. 91–94 ISBN 0444883908
- [18] Eckert, Roger; Randall, David (1983). *Animal physiology: mechanisms and adaptations*. San Francisco: W.H. Freeman. p. 239. ISBN 0-7167-1423-X.
- [19] López-Muñoz, F.; Boya, J., Alamo, C. (16 October 2006). "Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal". *Brain Research Bulletin* **70** (4–6): 391–405. doi:10.1016/j.brainresbull.2006.07.010. PMID 17027775.
- [20] Grant, Gunnar; Boya, J; Alamo, C (2007). "How the 1906 Nobel Prize in Physiology or Medicine was shared between Golgi and Cajal". *Brain Research Reviews* **55** (4–6): 490–8. doi:10.1016/j.brainresrev.2006.11.004. PMID 17306375.
- [21] Witcher M, Kirov S, Harris K (2007). "Plasticity of perisynaptic astroglia during synaptogenesis in the mature rat hippocampus". *Glia* **55** (1): 13–23. doi:10.1002/glia.20415. PMID 17001633.

- [22] Connors B, Long M (2004). "Electrical synapses in the mammalian brain". *Annu Rev Neurosci* **27**: 393–418. doi:10.1146/annurev.neuro.26.041002.131128. PMID 15217338.
- [23] Guillery, R. W. (2005). "Observations of synaptic structures: Origins of the neuron doctrine and its current status". *Philosophical Transactions of the Royal Society B: Biological Sciences* **360** (1458): 1281–1307. doi:10.1098/rstb.2003.1459. PMC 1569502.
- [24] Sabbatini R.M.E. April–July 2003. Neurons and Synapses: The History of Its Discovery ([http://www.cerebromente.org.br/n17/history/neurons3\\_i.htm](http://www.cerebromente.org.br/n17/history/neurons3_i.htm)). *Brain & Mind Magazine*, 17. Retrieved on March 19, 2007.
- [25] Djuricic M, Antic S, Chen W, Zecevic D (2004). "Voltage imaging from dendrites of mitral cells: EPSP attenuation and spike trigger zones". *J Neurosci* **24** (30): 6703–14. doi:10.1523/JNEUROSCI.0307-04.2004. PMID 15282273.
- [26] Cochilla, AJ; Alford, S (1997). "Glutamate receptor-mediated synaptic excitation in axons of the lamprey". *The Journal of physiology* **499** (Pt 2): 443–57. PMC 1159318. PMID 9080373.
- [27] Williams RW, Herrup K (1988). "The control of neuron number". *Annual Review of Neuroscience* **11**: 423–53. doi:10.1146/annurev.ne.11.030188.002231. PMID 3284447.
- [28] Azevedo FA, Carvalho LR, Grinberg LT, *et al.* (April 2009). "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain". *The Journal of Comparative Neurology* **513** (5): 532–41. doi:10.1002/cne.21974. PMID 19226510.
- [29] Wadep, Nicholas (1999-10-15). "Brain may grow new cells daily" (<http://www.nytimes.com/1999/10/15/us/brain-may-grow-new-cells-daily.html>). *Nytimes.com*. Retrieved 2013-02-16.
- [30] Ewen Callaway (26 May 2011). "How to make a human neuron" (<http://www.nature.com/news/2011/110526/full/news.2011.328.html>). *NatureNews*. Retrieved 2011-05-24. "By transforming cells from human skin into working nerve cells, researchers may have come up with a model for nervous-system diseases and perhaps even regenerative therapies based on cell transplants. The achievement, reported online today in *Nature*, is the latest in a fast-moving field called transdifferentiation, in which cells are forced to adopt new identities. In the past year, researchers have converted connective tissue cells found in skin into heart cells, blood cells and liver cells."

## Further reading

- Kandel E.R., Schwartz, J.H., Jessell, T.M. 2000. *Principles of Neural Science*, 4th ed., McGraw-Hill, New York.
- Bullock, T.H., Bennett, M.V.L., Johnston, D., Josephson, R., Marder, E., Fields R.D. 2005. *The Neuron Doctrine, Redux*, Science, V.310, p. 791–793.
- Ramón y Cajal, S. 1933 *Histology*, 10th ed., Wood, Baltimore.
- Richard S. Snell: *Clinical neuroanatomy* (Lippincott Williams & Wilkins, Ed.6th 2006) Philadelphia, Baltimore, New York, London. ISBN 978-963-226-293-2
- Roberts A., Bush B.M.H. 1981. *Neurons Without Impulses*. Cambridge University Press, Cambridge.
- Peters, A., Palay, S.L., Webster, H, D., 1991 *The Fine Structure of the Nervous System*, 3rd ed., Oxford, New York

## External links

- IBRO (International Brain Research Organization) (<http://www.ibro.info>). Fostering neuroscience research especially in less well-funded countries.
- NeuronBank (<http://NeuronBank.org>) an online neuromics tool for cataloging neuronal types and synaptic connectivity.
- High Resolution Neuroanatomical Images of Primate and Non-Primate Brains (<http://brainmaps.org>).
- The Department of Neuroscience at Wikiversity, which presently offers two courses: Fundamentals of Neuroscience and Comparative Neuroscience.
- NIF Search – Renshaw Cell ([http://www.neuinfo.org/nif/nifgw.html?query="Renshaw Cell"](http://www.neuinfo.org/nif/nifgw.html?query=)) via the Neuroscience Information Framework
- Cell Centered Database – Neuron (<http://ccdb.ucsd.edu/sand/main?event=showMPByType&typeid=0&start=1&pl=y>)
- Complete list of neuron types (<http://neurolex.org/wiki/Category:Neuron>) according to the Petilla convention, at NeuroLex.
- NeuroMorpho.Org (<http://NeuroMorpho.org>) an online database of digital reconstructions of neuronal morphology.

- Immunohistochemistry Image Gallery: Neuron ([http://www.immunoport.com/modules.php?name=gallery2&g2\\_view=keyalbum.KeywordAlbum&g2\\_keyword=Neuron](http://www.immunoport.com/modules.php?name=gallery2&g2_view=keyalbum.KeywordAlbum&g2_keyword=Neuron))

## Fractal

---

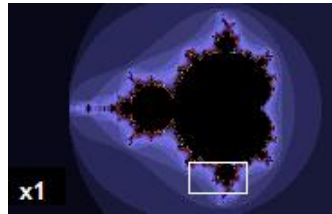


Figure 1a. The Mandelbrot set illustrates self-similarity. As you zoom in on the image at finer and finer scales, the same pattern re-appears so that it is virtually impossible to know at which level you are looking.



Figure 1b. Mandelbrot zoomed 6x



Figure 1c. Mandelbrot zoomed 100x

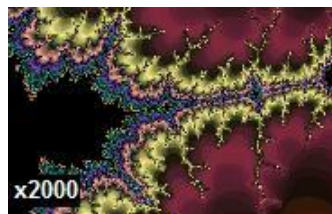


Figure 1d. Even 2000 times magnification of the Mandelbrot set uncovers fine detail resembling the full set.

A **fractal** is a mathematical set that has a fractal dimension that usually exceeds its topological dimension<sup>[1]</sup> and may fall between the integers.<sup>[2]</sup> Fractals are typically self-similar patterns, where self-similar means they are "the same from near as from far".<sup>[3]</sup> Fractals may be exactly the same at every scale, or, as illustrated in Figure 1, they may be nearly the same at different scales.<sup>[2][4][5][6]</sup> The definition of fractal goes beyond self-similarity *per se* to exclude trivial self-similarity and include the idea of a *detailed pattern* repeating itself.<sup>[2]:166; 18[4][7]</sup>

As mathematical equations, fractals are usually nowhere differentiable.<sup>[2][6][8]</sup> An infinite fractal curve can be perceived of as winding through space differently from an ordinary line, still being a 1-dimensional line yet having a fractal dimension indicating it also resembles a surface.<sup>[1]:48[2]:15</sup>

The mathematical roots of the idea of fractals have been traced through a formal path of published works, starting in the 17th century with notions of recursion, then moving through increasingly rigorous mathematical treatment of the concept to the study of continuous but not differentiable functions in the 19th century, and on to the coining of the word *fractal* in the 20th century with a subsequent burgeoning of interest in fractals and computer-based modelling

in the 21st century.<sup>[9][10]</sup> The term "fractal" was first used by mathematician Benoît Mandelbrot in 1975. Mandelbrot based it on the Latin *fractus* meaning "broken" or "fractured", and used it to extend the concept of theoretical fractional dimensions to geometric patterns in nature.<sup>[2]:405[7]</sup>

There is some disagreement amongst authorities about how the concept of a fractal should be formally defined. The general consensus is that theoretical fractals are infinitely self-similar, iterated, and detailed mathematical constructs having fractal dimensions, of which many examples have been formulated and studied in great depth.<sup>[2][4][5]</sup> Fractals are not limited to geometric patterns, but can also describe processes in time.<sup>[3][6][11]</sup> Fractal patterns with various degrees of self-similarity have been rendered or studied in images, structures and sounds<sup>[12]</sup> and found in nature,<sup>[13][14][15][16][17]</sup> technology,<sup>[18][19][20][21]</sup> art,<sup>[22][23][24]</sup> and law.<sup>[25]</sup>

## Introduction

The word "fractal" often has different connotations for laypeople than mathematicians, where the layperson is more likely to be familiar with fractal art than a mathematical conception. The mathematical concept is difficult to formally define even for mathematicians, but key features can be understood with little mathematical background.

The feature of "self-similarity", for instance, is easily understood by analogy to zooming in with a lens or other device that zooms in on digital images to uncover finer, previously invisible, new structure. If this is done on fractals, however, no new detail appears; nothing changes and the same pattern repeats over and over, or for some fractals, nearly the same pattern reappears over and over. Self-similarity itself is not necessarily counter-intuitive (e.g., people have pondered self-similarity informally such as in the infinite regress in parallel mirrors or the homunculus, the little man inside the head of the little man inside the head...). The difference for fractals is that the pattern reproduced must be **detailed**.<sup>[2]:166; 18[4][7]</sup>

This idea of being detailed relates to another feature that can be understood without mathematical background: Having a fractional or fractal dimension greater than its topological dimension, for instance, refers to how a fractal scales compared to how geometric shapes are usually perceived. A regular line, for instance, is conventionally understood to be 1-dimensional; if such a curve is divided into pieces each 1/3 the length of the original, there are always 3 equal pieces. In contrast, consider the curve in Figure 2. It is also 1-dimensional for the same reason as the ordinary line, but it has, in addition, a fractal dimension greater than 1 because of how its detail can be measured. The fractal curve divided into parts 1/3 the length of the original line becomes 4 pieces rearranged to repeat the original detail, and this unusual relationship is the basis of its fractal dimension.

This also leads to understanding a third feature, that fractals as mathematical equations are "nowhere differentiable". In a concrete sense, this means fractals cannot be measured in traditional ways.<sup>[2][6][8]</sup> To elaborate, in trying to find the length of a wavy non-fractal curve, one could find straight segments of some measuring tool small enough to lay end to end over the waves, where the pieces could get small enough to be considered to conform to the curve in the normal manner of measuring with a tape measure. But in measuring a wavy **fractal** curve such as the one in Figure 2, one would never find a small enough straight segment to conform to the curve, because the wavy pattern would always re-appear, albeit at a smaller size, essentially pulling a little more of the tape measure into the total length measured each time one attempted to fit it tighter and tighter to the curve. This is perhaps counter-intuitive, but it is how fractals behave.<sup>[2]</sup>

## History

The history of fractals traces a path from chiefly theoretical studies to modern applications in computer graphics, with several notable people contributing canonical fractal forms along the way.<sup>[9][10]</sup> According to Pickover, the mathematics behind fractals began to take shape in the 17th century when the mathematician and philosopher Gottfried Leibniz pondered recursive self-similarity (although he made the mistake of thinking that only the straight line was self-similar in this sense).<sup>[26]</sup> In his writings, Leibniz used the term "fractional exponents", but lamented that "Geometry" did not yet know of them.<sup>[2]:405</sup> Indeed, according to various historical accounts, after that point few mathematicians tackled the issues and the work of those who did remained obscured largely because of resistance to such unfamiliar emerging concepts, which were sometimes referred to as mathematical "monsters".<sup>[8][9][10]</sup> Thus, it was not until two centuries had passed that in 1872 Karl Weierstrass presented the first definition of a function with a graph that would today be considered fractal, having the non-intuitive property of being everywhere continuous but nowhere differentiable.<sup>[9]:7[10]</sup> Not long after that, in 1883, Georg Cantor, who attended lectures by Weierstrass,<sup>[10]</sup> published examples of subsets of the real line known as Cantor sets, which had unusual properties and are now recognized as fractals.<sup>[9]:11-24</sup> Also in the last part of that century, Felix Klein and Henri Poincaré introduced a category of fractal that has come to be called "self-inverse" fractals.<sup>[2]:166</sup>

One of the next milestones came in 1904, when Helge von Koch, extending ideas of Poincaré and dissatisfied with Weierstrass's abstract and analytic definition, gave a more geometric definition including hand drawn images of a similar function, which is now called the Koch curve (see Figure 2)<sup>[9]:25 [10]</sup> Another milestone came a decade later in 1915, when Waclaw Sierpiński constructed his famous triangle then, one year later, his carpet. By 1918, two French mathematicians, Pierre Fatou and Gaston Julia, though working independently, arrived essentially simultaneously at results describing what are now seen as fractal behaviour associated with mapping complex numbers and iterative functions and leading to further ideas about attractors and repellers (i.e., points that attract or repel other points), which have become very important in the study of fractals (see Figure 3 and Figure 4).<sup>[6][9][10]</sup> Very shortly after that work was submitted, by March 1918, Felix Hausdorff expanded the definition of "dimension", significantly for the evolution of the definition of fractals, to allow for sets to have noninteger dimensions.<sup>[10]</sup> The idea of self-similar curves was taken further by Paul Pierre Lévy, who, in his 1938 paper *Plane or Space Curves and Surfaces Consisting of Parts Similar to the Whole* described a new fractal curve, the Lévy C curve.<sup>[27]</sup>

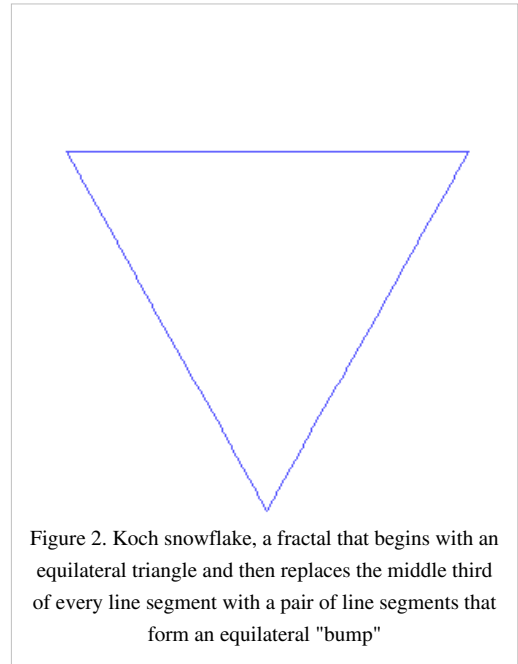


Figure 2. Koch snowflake, a fractal that begins with an equilateral triangle and then replaces the middle third of every line segment with a pair of line segments that form an equilateral "bump"

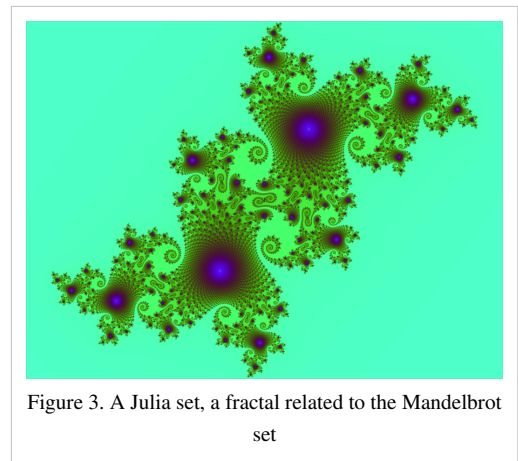


Figure 3. A Julia set, a fractal related to the Mandelbrot set

Different researchers have postulated that without the aid of modern computer graphics, early investigators were limited to what they could depict in manual drawings, so lacked the means to visualize the beauty and appreciate some of the implications of many of the patterns they had discovered (the Julia set, for instance, could only be visualized through a few iterations as very simple drawings hardly resembling the image in Figure 3).<sup>[2]:179[8][10]</sup> That changed, however, in the 1960s, when Benoît Mandelbrot started writing about self-similarity in papers such as *How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension*,<sup>[28]</sup> which built on earlier work by Lewis Fry Richardson. In 1975<sup>[7]</sup> Mandelbrot solidified hundreds of years of thought and mathematical development in coining the word "fractal" and illustrated his mathematical definition with striking computer-constructed visualizations. These images, such as of his canonical Mandelbrot set pictured in Figure 1, captured the popular imagination; many of them were based on recursion, leading to the popular meaning of the term "fractal".<sup>[29]</sup> Currently, fractal studies are essentially exclusively computer-based.<sup>[8][9][26]</sup>

## Characteristics

One often cited description that Mandelbrot published to describe geometric fractals is "a rough or fragmented geometric shape that can be split into parts, each of which is (at least approximately) a reduced-size copy of the whole";<sup>[2]</sup> this is generally helpful but limited. Authorities disagree on the exact definition of *fractal*, but most usually elaborate on the basic ideas of self-similarity and an unusual relationship with the space a fractal is embedded in.<sup>[2][3][4] [6][30]</sup> One point agreed on is that fractal patterns are characterized by fractal dimensions, but whereas these numbers quantify complexity (i.e., changing detail with changing scale), they neither uniquely describe nor specify details of how to construct particular fractal patterns.<sup>[31]</sup> In 1975 when Mandelbrot coined the word "fractal", he did so to denote an object whose Hausdorff–Besicovitch dimension is greater than its topological dimension.<sup>[7]</sup> It has been noted that this dimensional requirement is not met by fractal space-filling curves such as the Hilbert curve.<sup>[32]</sup>

According to Falconer, rather than being strictly defined, fractals should, in addition to being nowhere differentiable and able to have a fractal dimension, be generally characterized by a gestalt of the following features;<sup>[4]</sup>

- Self-similarity, which may be manifested as:
  - Exact self-similarity: *identical at all scales; e.g. Koch snowflake*
  - Quasi self-similarity: *approximates the same pattern at different scales; may contain small copies of the entire fractal in distorted and degenerate forms; e.g., the Mandelbrot set's satellites are approximations of the entire set, but not exact copies, as shown in Figure 1*
  - Statistical self-similarity: *repeats a pattern stochastically so numerical or statistical measures are preserved across scales; e.g., randomly generated fractals; the well-known example of the coastline of Britain, for which one would not expect to find a segment scaled and repeated as neatly as the repeated unit*

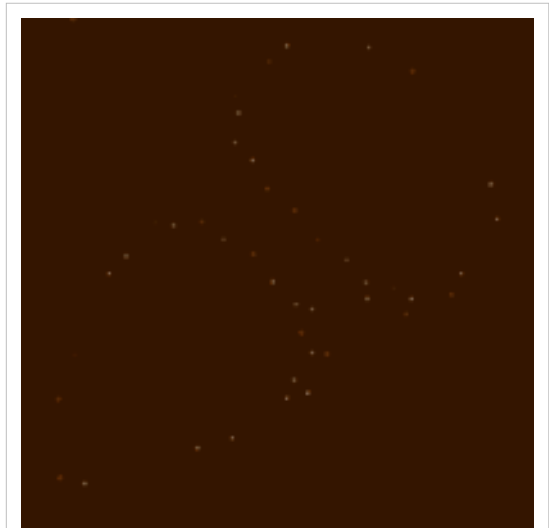
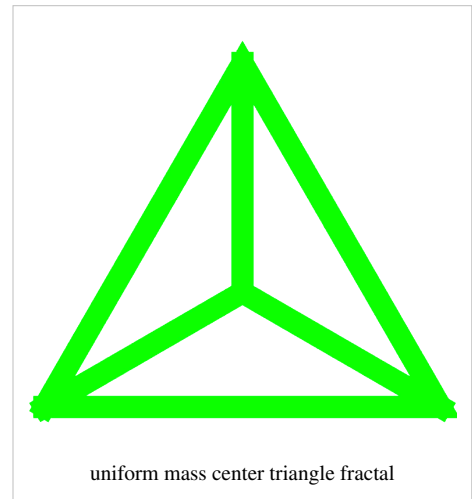


Figure 4. A strange attractor that exhibits multifractal scaling



uniform mass center triangle fractal

that defines, for example, the Koch snowflake<sup>[6]</sup>

- Qualitative self-similarity: *as in a time series*<sup>[11]</sup>
- Multifractal scaling: *characterized by more than one fractal dimension or scaling rule*
- Fine or detailed structure at arbitrarily small scales. A consequence of this structure is fractals may have emergent properties<sup>[33]</sup> (related to the next criterion in this list).
- Irregularity locally and globally that is not easily described in traditional Euclidean geometric language. For images of fractal patterns, this has been expressed by phrases such as "smoothly piling up surfaces" and "swirls upon swirls".<sup>[1]</sup>
- Simple and "perhaps recursive" definitions *see Common techniques for generating fractals*

As a group, these criteria form guidelines for excluding certain cases, such as those that may be self-similar without having other typically fractal features. A straight line, for instance, is self-similar but not fractal because it lacks detail, is easily described in Euclidean language, has the same Hausdorff dimension as topological dimension, and is fully defined without a need for recursion.<sup>[2][6]</sup>

## Common techniques for generating fractals

Images of fractals can be created by fractal generating programs.

- *Iterated function systems* – use fixed geometric replacement rules; may be stochastic or deterministic;<sup>[34]</sup> e.g., Koch snowflake, Cantor set, Haferman carpet,<sup>[35]</sup> Sierpinski carpet, Sierpinski gasket, Peano curve, Harter-Heighway dragon curve, T-Square, Menger sponge
- *Strange attractors* – use iterations of a map or solutions of a system of initial-value differential equations that exhibit chaos (e.g., see multifractal image)
- *L-systems* - use string rewriting; may resemble branching patterns, such as in plants, biological cells (e.g., neurons and immune system cells<sup>[17]</sup>), blood vessels, pulmonary structure,<sup>[36]</sup> etc. (e.g., see Figure 5) or turtle graphics patterns such as space-filling curves and tilings
- *Escape-time fractals* – use a formula or recurrence relation at each point in a space (such as the complex plane); usually quasi-self-similar; also known as "orbit" fractals; e.g., the Mandelbrot set, Julia set, Burning Ship fractal, Nova fractal and Lyapunov fractal. The 2d vector fields that are generated by one or two iterations of escape-time formulae also give rise to a fractal form when points (or pixel data) are passed through this field repeatedly.
- *Random fractals* – use stochastic rules; e.g., Lévy flight, percolation clusters, self avoiding walks, fractal landscapes, trajectories of Brownian motion and the Brownian tree (i.e., dendritic fractals generated by modeling diffusion-limited aggregation or reaction-limited aggregation clusters).<sup>[6]</sup>

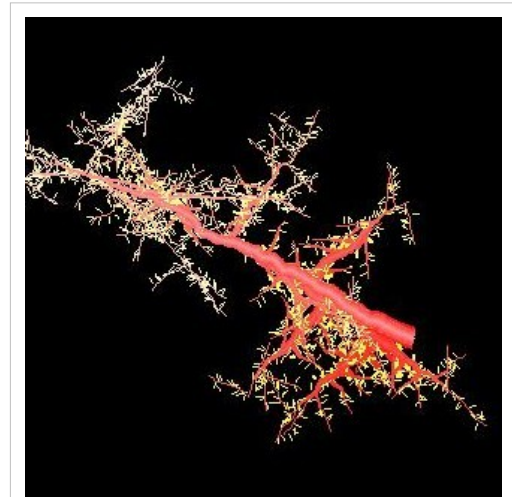
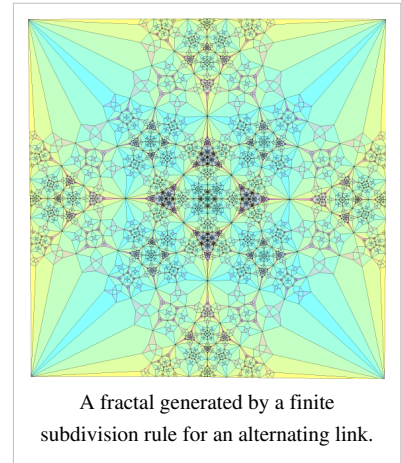


Figure 5. Self-similar branching pattern modeled in silico using L-systems principles<sup>[17]</sup>

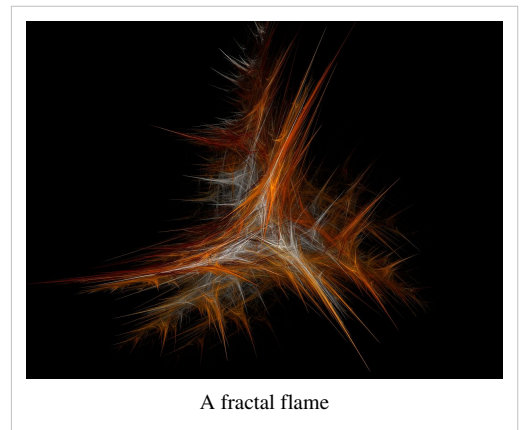


- *Finite subdivision rules* use a recursive topological algorithm for refining tilings<sup>[37]</sup> and they are similar to the process of cell division.<sup>[38]</sup> The iterative processes used in creating the Cantor set and the Sierpinski carpet are examples of finite subdivision rules, as is barycentric subdivision.



## Simulated fractals

Fractal patterns have been modeled extensively, albeit within a range of scales rather than infinitely, owing to the practical limits of physical time and space. Models may simulate theoretical fractals or natural phenomena with fractal features. The outputs of the modelling process may be highly artistic renderings, outputs for investigation, or benchmarks for fractal analysis. Some specific applications of fractals to technology are listed elsewhere. Images and other outputs of modelling are normally referred to as being "fractals" even if they do not have strictly fractal characteristics, such as when it is possible to zoom into a region of the fractal image that does not exhibit any fractal properties. Also, these may include calculation or display artifacts which are not characteristics of true fractals.



Modeled fractals may be sounds,<sup>[12]</sup> digital images, electrochemical patterns, circadian rhythms,<sup>[39]</sup> etc. Fractal patterns have been reconstructed in physical 3-dimensional space<sup>[20]:10</sup> and virtually, often called "in silico" modeling.<sup>[36]</sup> Models of fractals are generally created using fractal-generating software that implements techniques such as those outlined above.<sup>[6][11][20]</sup> As one illustration, trees, ferns, cells of the nervous system,<sup>[17]</sup> blood and lung vasculature,<sup>[36]</sup> and other branching patterns in nature can be modeled on a computer by using recursive algorithms and L-systems techniques.<sup>[17]</sup> The recursive nature of some patterns is obvious in certain examples—a branch from a tree or a frond from a fern is a miniature replica of the whole: not identical, but similar in nature. Similarly, random fractals have been used to describe/create many highly irregular real-world objects. A limitation of modeling fractals is that resemblance of a fractal model to a natural phenomenon does not prove that the phenomenon being modeled is formed by a process similar to the modeling algorithm.

## Natural phenomena with fractal features

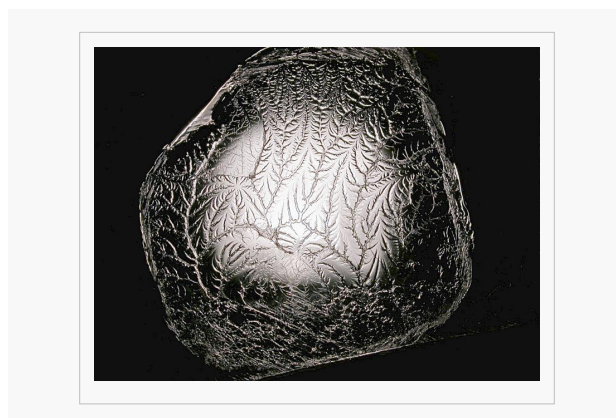
Approximate fractals found in nature display self-similarity over extended, but finite, scale ranges. The connection between fractals and leaves, for instance, is currently being used to determine how much carbon is contained in trees.<sup>[40]</sup>

Examples of phenomena known or anticipated to have fractal features are listed below:

<ul style="list-style-type: none"> <li>• clouds</li> <li>• river networks</li> <li>• fault lines</li> <li>• mountain ranges</li> <li>• craters</li> <li>• lightning bolts</li> <li>• coastlines</li> <li>• Mountain Goat horns</li> </ul>	<ul style="list-style-type: none"> <li>• animal coloration patterns</li> <li>• Romanesco broccoli</li> <li>• heart rates<sup>[13]</sup></li> <li>• heartbeat<sup>[14]</sup></li> <li>• earthquakes<sup>[21][43]</sup></li> <li>• snow flakes<sup>[44]</sup></li> </ul>	<ul style="list-style-type: none"> <li>• crystals<sup>[41]</sup></li> <li>• blood vessels and pulmonary vessels<sup>[36]</sup></li> <li>• ocean waves<sup>[42]</sup></li> <li>• DNA</li> <li>• various vegetables (cauliflower &amp; broccoli)</li> <li>• Psychological subjective perception<sup>[45]</sup></li> </ul>
---	--	---



Frost crystals formed naturally on cold glass illustrate fractal process development in a purely physical system



A fractal is formed when pulling apart two glue-covered acrylic sheets



High voltage breakdown within a 4" block of acrylic creates a fractal Lichtenberg figure



Romanesco broccoli, showing self-similar form approximating a natural fractal

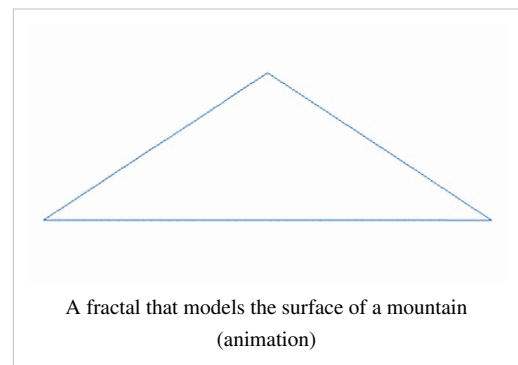
## In creative works

Fractal patterns have been found in the paintings of American artist Jackson Pollock. While Pollock's paintings appear to be composed of chaotic dripping and splattering, computer analysis has found fractal patterns in his work.<sup>[24]</sup>

Decalcomania, a technique used by artists such as Max Ernst, can produce fractal-like patterns.<sup>[46]</sup> It involves pressing paint between two surfaces and pulling them apart.

Cyberneticist Ron Eglash has suggested that fractal geometry and mathematics are prevalent in African art, games, divination, trade, and architecture. Circular houses appear in circles of circles, rectangular houses in rectangles of rectangles, and so on. Such scaling patterns can also be found in African textiles, sculpture, and even cornrow hairstyles.<sup>[23][47]</sup>

In a 1996 interview with Michael Silverblatt, David Foster Wallace admitted that the structure of the first draft of *Infinite Jest* he gave to his editor Michael Pietsch was inspired by fractals, specifically the Sierpinski triangle (aka Sierpinski gasket) but that the edited novel is "more like a lopsided Sierpinsky Gasket".<sup>[22]</sup>



## In law

If a rule or principle of law is conceptualized as defining a two-dimensional "area" of conduct, conduct within which should be legal and conduct outside of which should be illegal, it has been observed that the border of that area must be a fractal, because of the infinite and recursive potential exceptions and extensions necessary to account appropriately for all variations in fact pattern that may arise.<sup>[25]</sup>

## Applications in technology

<ul style="list-style-type: none"> <li>• fractal antennas<sup>[48]</sup></li> <li>• Fractal transistor<sup>[49]</sup></li> <li>• fractal heat exchangers</li> <li>• digital imaging</li> <li>• urban growth<sup>[52][53]</sup></li> <li>• Classification of histopathology slides</li> <li>• Fractal landscape or Coastline complexity</li> <li>• Enzyme/enzymology (Michaelis-Menten kinetics)</li> <li>• Generation of new music</li> <li>• Signal and image compression</li> <li>• Creation of digital photographic enlargements</li> <li>• Fractal in soil mechanics</li> </ul>	<ul style="list-style-type: none"> <li>• Computer and video game design</li> <li>• computer graphics</li> <li>• organic environments</li> <li>• procedural generation</li> <li>• Fractography and fracture mechanics</li> <li>• Small angle scattering theory of fractally rough systems</li> <li>• T-shirts and other fashion</li> <li>• Generation of patterns for camouflage, such as MARPAT</li> <li>• Digital sundial</li> <li>• Technical analysis of price series</li> <li>• Fractals in networks</li> </ul>	<ul style="list-style-type: none"> <li>• medicine<sup>[20]</sup></li> <li>• neuroscience<sup>[15][16]</sup></li> <li>• diagnostic imaging<sup>[19]</sup></li> <li>• pathology<sup>[50][51]</sup></li> <li>• geology<sup>[54]</sup></li> <li>• geography<sup>[55]</sup></li> <li>• archaeology<sup>[56][57]</sup></li> <li>• soil mechanics<sup>[18]</sup></li> <li>• seismology<sup>[21]</sup></li> <li>• search and rescue<sup>[58]</sup></li> <li>• technical analysis<sup>[59]</sup></li> </ul>
---	---	--

## Notes

- [1] Mandelbrot, Benoît B. (2004). *Fractals and Chaos*. Berlin: Springer. p. 38. ISBN 978-0-387-20158-0. "A fractal set is one for which the fractal (Hausdorff-Besicovitch) dimension strictly exceeds the topological dimension"
- [2] Mandelbrot, Benoît B. (1983). *The fractal geometry of nature* (<http://books.google.com/books?id=0R2LkE3N7-oC>). Macmillan. ISBN 978-0-7167-1186-5. . Retrieved 1 February 2012.
- [3] Gouyet, Jean-François (1996). *Physics and fractal structures*. Paris/New York: Masson Springer. ISBN 978-0-387-94153-0.
- [4] Falconer, Kenneth (2003). *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley & Sons, Ltd.. xxv. ISBN 0-470-84862-6.
- [5] Briggs, John (1992). *Fractals: The Patterns of Chaos*. London, UK: Thames and Hudson. p. 148. ISBN 0-500-27693-5, 0500276935.
- [6] Vicsek, Tamás (1992). *Fractal growth phenomena*. Singapore/New Jersey: World Scientific. pp. 31; 139–146. ISBN 978-981-02-0668-0.
- [7] Albers, Donald J.; Alexanderson, Gerald L. (2008). "Benoît Mandelbrot: In his own words". *Mathematical people : profiles and interviews*. Wellesley, MA: AK Peters. p. 214. ISBN 978-1-56881-340-0.
- [8] Gordon, Nigel (2000). *Introducing fractal geometry*. Duxford, UK: Icon. p. 71. ISBN 978-1-84046-123-7.
- [9] Edgar, Gerald (2004). *Classics on Fractals*. Boulder, CO: Westview Press. ISBN 978-0-8133-4153-8.
- [10] Trochet, Holly (2009). "A History of Fractal Geometry" (<http://www.webcitation.org/65DCT2znx>). *MacTutor History of Mathematics*. Archived from the original on 4 February 2012. .
- [11] Peters, Edgar (1996). *Chaos and order in the capital markets : a new view of cycles, prices, and market volatility*. New York: Wiley. ISBN 0-471-13938-6.
- [12] Brothers, Harlan J. (2007). "Structural Scaling in Bach's Cello Suite No. 3". *Fractals* **15**: 89–95. doi:10.1142/S0218348X0700337X.
- [13] Tan, Can Ozan; Cohen, Michael A.; Eckberg, Dwain L.; Taylor, J. Andrew (2009). "Fractal properties of human heart period variability: Physiological and methodological implications". *The Journal of Physiology* **587** (15): 3929. doi:10.1113/jphysiol.2009.169219.
- [14] Buldyrev, Sergey V.; Goldberger, Ary L.; Havlin, Shlomo; Peng, Chung-Kang; Stanley, H. Eugene (1995). "3" ([http://havlin.biu.ac.il/ShlomoHavlinbooks\\_f\\_in\\_s.php](http://havlin.biu.ac.il/ShlomoHavlinbooks_f_in_s.php)). In Bunde, Armin; Havlin, Shlomo. *Fractals in Science*. Springer. .
- [15] Liu, Jing Z.; Zhang, Lu D.; Yue, Guang H. (2003). "Fractal Dimension in Human Cerebellum Measured by Magnetic Resonance Imaging". *Biophysical Journal* **85** (6): 4041–4046. doi:10.1016/S0006-3495(03)74817-6. PMC 1303704. PMID 14645092.
- [16] Karperien, Audrey L.; Jelinek, Herbert F.; Buchan, Alastair M. (2008). "Box-Counting Analysis of Microglia Form in Schizophrenia, Alzheimer's Disease and Affective Disorder". *Fractals* **16** (2): 103. doi:10.1142/S0218348X08003880.

- [17] Jelinek, Herbert F.; Karperien, Audrey; Cornforth, David; Cesar, Roberto; Leandro, Jorge de Jesus Gomes (2002). "MicroMod-an L-systems approach to neural modelling" (<http://books.google.com/books?id=FFSUGQAACA AJ>). In Sarker, Ruhul. *Workshop proceedings: the Sixth Australia-Japan Joint Workshop on Intelligent and Evolutionary Systems, University House, ANU.*, University of New South Wales. ISBN 9780731705054. OCLC 224846454. [http://researchoutput.csu.edu.au/R/-?func=dbin-jump-full&object\\_id=6595&local\\_base=GEN01-CSU01](http://researchoutput.csu.edu.au/R/-?func=dbin-jump-full&object_id=6595&local_base=GEN01-CSU01). Retrieved 3 February 2012.
- [18] Hu, Shougeng; Cheng, Qiuming; Wang, Le; Xie, Shuyun (2012). "Multifractal characterization of urban residential land price in space and time". *Applied Geography* **34**: 161. doi:10.1016/j.apgeog.2011.10.016.
- [19] Karperien, Audrey; Jelinek, Herbert F.; Leandro, Jorge de Jesus Gomes; Soares, João V. B.; Cesar Jr, Roberto M.; Luckie, Alan (2008). "Automated detection of proliferative retinopathy in clinical practice". *Clinical ophthalmology (Auckland, N.Z.)* **2** (1): 109–122. doi:10.2147/OPHTH.S1579. PMC 2698675. PMID 19668394.
- [20] Losa, Gabriele A.; Nonnenmacher, Theo F. (2005). *Fractals in biology and medicine* (<http://books.google.com/books?id=t919GdAt95gC>). Springer. ISBN 978-3-7643-7172-2. Retrieved 1 February 2012.
- [21] Vannucchi, Paola; Leoni, Lorenzo (2007). "Structural characterization of the Costa Rica décollement: Evidence for seismically-induced fluid pulsing". *Earth and Planetary Science Letters* **262** (3–4): 413. Bibcode 2007E&PSL.262..413V. doi:10.1016/j.epsl.2007.07.056.
- [22] Wallace, David Foster. "Bookworm on KCRW" ([http://www.kcrw.com/etc/programs/bw/bw960411david\\_foster\\_wallace](http://www.kcrw.com/etc/programs/bw/bw960411david_foster_wallace)). Kcrw.com. Retrieved 2010-10-17.
- [23] Eglash, Ron (1999). "African Fractals: Modern Computing and Indigenous Design" (<http://www.rpi.edu/~eglash/eglash.dir/afactal/afactal.htm>). New Brunswick: Rutgers University Press. Retrieved 2010-10-17.
- [24] Taylor, Richard; Micolich, Adam P.; Jonas, David. "Fractal Expressionism: Can Science Be Used To Further Our Understanding Of Art?" ([http://www.phys.unsw.edu.au/PHYSICS\\_/FRACTAL\\_EXPRESSIONISM/fractal\\_taylor.html](http://www.phys.unsw.edu.au/PHYSICS_/FRACTAL_EXPRESSIONISM/fractal_taylor.html)). Phys.unsw.edu.au. Retrieved 2010-10-17.
- [25] Stumpff, Andrew. "The Law is a Fractal: The Attempt to Anticipate Everything" ([http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2157804](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2157804)). Loyola University Chicago Law Journal. Retrieved 2012-10-05.
- [26] Pickover, Clifford A. (2009). *The Math Book: From Pythagoras to the 57th Dimension, 250 Milestones in the History of Mathematics* (<http://books.google.com/?id=JrsIMKTgSZwC&pg=PA310&dq=fractal+koch+curve+book#v=onepage&q=fractal+koch+curve+book&f=false>). Sterling Publishing Company, Inc.. p. 310. ISBN 978-1-4027-5796-9. Retrieved 2011-02-05.
- [27] The original paper, Lévy, Paul (1938). "Les Courbes planes ou gauches et les surfaces composées de parties semblables au tout". *Journal de l'École Polytechnique*: 227–247, 249–291., is translated in Edgar, pages 181-239.
- [28] Batty, Michael (1985-04-04). "Fractals - Geometry Between Dimensions" ([http://books.google.com/?id=sz6GAq\\_PsmMC&pg=PA31&dq=mandelbrot+how+long+is+the+coast+of+Britain+Science+book#v=onepage&q=mandelbrot+how+long+is+the+coast+of+Britain+Science+book&f=false](http://books.google.com/?id=sz6GAq_PsmMC&pg=PA31&dq=mandelbrot+how+long+is+the+coast+of+Britain+Science+book#v=onepage&q=mandelbrot+how+long+is+the+coast+of+Britain+Science+book&f=false)). *New Scientist* (Holborn Publishing Group) **105** (1450): 31. .
- [29] Russ, John C. (1994). *Fractal surfaces* (<http://books.google.com/?id=qDQjyuuDRxUC&pg=PA1&lpg=PA1&dq=fractal+history+book#v=onepage&q=fractal+history+book&f=false>). **1**. Springer. p. 1. ISBN 978-0-306-44702-0. Retrieved 2011-02-05.
- [30] Edgar, Gerald (2008). *Measure, topology, and fractal geometry*. New York, NY: Springer-Verlag. p. 1. ISBN 978-0-387-74748-4.
- [31] Karperien, Audrey (2004). <http://www.webcitation.org/65DyLbmF1> *Defining microglial morphology: Form, Function, and Fractal Dimension* ([http://bilby.unilinc.edu.au/view/action/singleViewer.do?dvs=1328442053986~182&locale=en\\_US&VIEWER\\_URL=/view/action/singleViewer.do?&DELIVERY\\_RULE\\_ID=10&search\\_terms=karperien&adjacency=N&application=DIGITool-3&frameId=1&usePid1=true&usePid2=true](http://bilby.unilinc.edu.au/view/action/singleViewer.do?dvs=1328442053986~182&locale=en_US&VIEWER_URL=/view/action/singleViewer.do?&DELIVERY_RULE_ID=10&search_terms=karperien&adjacency=N&application=DIGITool-3&frameId=1&usePid1=true&usePid2=true),). Charles Sturt University. Retrieved 2012-02-05.
- [32] The Hilbert curve map is not a homeomorphism, so it does not preserve topological dimension. The topological dimension and Hausdorff dimension of the image of the Hilbert map in  $\mathbf{R}^2$  are both 2. Note, however, that the topological dimension of the *graph* of the Hilbert map (a set in  $\mathbf{R}^3$ ) is 1.
- [33] Spencer, John; Thomas, Michael S. C.; McClelland, James L. (2009). *Toward a unified theory of development : connectionism and dynamic systems theory re-considered*. Oxford/New York: Oxford University Press. ISBN 978-0-19-530059-8.
- [34] Frame, Angus (3 August 1998). "Iterated Function Systems" (<http://books.google.com/books?id=A51ARsapVuUC>). In Pickover, Clifford A.. *Chaos and fractals: a computer graphical journey : ten year compilation of advanced research*. Elsevier. pp. 349–351. ISBN 978-0-444-50002-1. Retrieved 4 February 2012.
- [35] "Haferman Carpet" (<http://www.wolframalpha.com/input/?i=Haferman+carpet>). WolframAlpha. Retrieved 18 October 2012.
- [36] Hahn, Horst K.; Georg, Manfred; Peitgen, Heinz-Otto (2005). "Fractal aspects of three-dimensional vascular constructive optimization" (<http://books.google.com/books?id=t919GdAt95gC>). In Losa, Gabriele A.; Nonnenmacher, Theo F.. *Fractals in biology and medicine*. Springer. pp. 55–66. ISBN 978-3-7643-7172-2. >
- [37] J. W. Cannon, W. J. Floyd, W. R. Parry. *Finite subdivision rules*. Conformal Geometry and Dynamics, vol. 5 (2001), pp. 153–196.
- [38] J. W. Cannon, W. Floyd and W. Parry. *Crystal growth, biological cell growth and geometry*. ([http://books.google.co.uk/books?id=qZHqyUli9y8C&pg=PA65&lpg=PA65&dq=james+w.+cannon+maths&source=web&ots=RP1svsBqga&sig=kxyEXFBqOG5NnJncng9HHniHyrc&hl=en&sa=X&oi=book\\_result&resnum=1&ct=result#PPA65,M1](http://books.google.co.uk/books?id=qZHqyUli9y8C&pg=PA65&lpg=PA65&dq=james+w.+cannon+maths&source=web&ots=RP1svsBqga&sig=kxyEXFBqOG5NnJncng9HHniHyrc&hl=en&sa=X&oi=book_result&resnum=1&ct=result#PPA65,M1)) Pattern Formation in Biology, Vision and Dynamics, pp. 65–82. World Scientific, 2000. ISBN 981-02-3792-8, ISBN 978-981-02-3792-9.
- [39] Fathallah-Shaykh, Hassan M. (2011). "Fractal Dimension of the Drosophila Circadian Clock". *Fractals* **19** (4): 423–430. doi:10.1142/S0218348X11005476.
- [40] "Hunting the Hidden Dimension." *Nova*. PBS. WPMB-Maryland. 28 October 2008.

- [41] Carbone, Alessandra; Gromov, Mikhael; Prusinkiewicz, Przemyslaw (2000). *Pattern formation in biology, vision and dynamics* (<http://books.google.com/?id=qZHyqUli9y8C&pg=PA78&dq=crystal+fractals+book#v=onepage&q=crystal+fractals+book&f=false>). World Scientific. p. 78. ISBN 978-981-02-3792-9. .
- [42] Addison, Paul S. (1997). *Fractals and chaos: an illustrated course* (<http://books.google.com/?id=l2E4ciBQ9qEC&pg=PA45&dq=lightning+fractals+book#v=onepage&q=lightning+fractals+book&f=false>). CRC Press. pp. 44–46. ISBN 978-0-7503-0400-9. . Retrieved 2011-02-05.
- [43] Sornette, Didier (2004). *Critical phenomena in natural sciences: chaos, fractals, selforganization, and disorder : concepts and tools*. Springer. pp. 128–140. ISBN 978-3-540-40754-6.
- [44] Meyer, Yves; Roques, Sylvie (1993). *Progress in wavelet analysis and applications: proceedings of the International Conference "Wavelets and Applications," Toulouse, France - June 1992* (<http://books.google.com/?id=aHux78oQbbkC&pg=PA25&dq=snowflake+fractals+book#v=onepage&q=snowflake+fractals+book&f=false>). Atlantica Séguier Frontières. p. 25. ISBN 978-2-86332-130-0. . Retrieved 2011-02-05.
- [45] Pincus, David (September 2009). "The Chaotic Life: Fractal Brains Fractal Thoughts" (<http://www.psychologytoday.com/blog/the-chaotic-life/200909/fractal-brains-fractal-thoughts>). *psychologytoday.com*. .
- [46] Frame, Michael; and Mandelbrot, Benoît B.; *A Panorama of Fractals and Their Uses* (<http://classes.yale.edu/Fractals/Panorama/>)
- [47] Nelson, Bryn; *Sophisticated Mathematics Behind African Village Designs Fractal patterns use repetition on large, small scale* (<http://www.sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive/2000/02/23/MN36684.DTL>), San Francisco Chronicle, Wednesday, February 23, 2009
- [48] Hohlfield, Robert G.; Cohen, Nathan (1999). "Self-similarity and the geometric requirements for frequency independence in Antennae". *Fractals* **7** (1): 79–84. doi:10.1142/S0218348X99000098.
- [49] Reiner, Richard; Waltereit, Patrick; Benkhelifa, Fouad; Müller, Stefan; Walcher, Herbert; Wagner, Sandrine; Quay, Rüdiger; Schlechtweg, Michael et al. (2012). "Fractal structures for low-resistance large area AlGaN/GaN power transistors" (<http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6229091&url=http://ieeexplore.ieee.org/iel5/6222389/6228999/06229091.pdf?arnumber=6229091>). *Proceedings of ISPSD*: 341. doi:10.1109/ISPSD.2012.6229091. ISBN 978-1-4577-1596-9. .
- [50] Smith, Robert F.; Mohr, David N.; Torres, Vicente E.; Offord, Kenneth P.; Melton III, L. Joseph (1989). "Renal insufficiency in community patients with mild asymptomatic microhematuria". *Mayo Clinic proceedings. Mayo Clinic* **64** (4): 409–414. PMID 2716356.
- [51] Landini, Gabriel (2011). "Fractals in microscopy". *Journal of Microscopy* **241** (1): 1–8. doi:10.1111/j.1365-2818.2010.03454.x. PMID 21118245.
- [52] Chen, Yanguang (2011). "Modeling Fractal Structure of City-Size Distributions Using Correlation Functions". *PLoS ONE* **6** (9): e24791. doi:10.1371/journal.pone.0024791. PMC 3176775. PMID 21949753.
- [53] "Applications" (<http://library.thinkquest.org/26242/full/ap/ap.html>). . Retrieved 2007-10-21.
- [54] Cheng, Qiuming (1997). "Multifractal Modeling and Lacunarity Analysis". *Mathematical Geology* **29** (7): 919–932. doi:10.1023/A:1022355723781.
- [55] Chen, Yanguang (2011). "Modeling Fractal Structure of City-Size Distributions Using Correlation Functions". *PLoS ONE* **6** (9): e24791. doi:10.1371/journal.pone.0024791. PMC 3176775. PMID 21949753.
- [56] Burkle-Elizondo, Gerardo; Valdéz-Cepeda, Ricardo David (2006). "Fractal analysis of Mesoamerican pyramids". *Nonlinear dynamics, psychology, and life sciences* **10** (1): 105–122. PMID 16393505.
- [57] Brown, Clifford T.; Witschey, Walter R. T.; Liebovitch, Larry S. (2005). "The Broken Past: Fractals in Archaeology". *Journal of Archaeological Method and Theory* **12**: 37. doi:10.1007/s10816-005-2396-6.
- [58] Saeedi, Panteha; Sorensen, Soren A.. "An Algorithmic Approach to Generate After-disaster Test Fields for Search and Rescue Agents" ([http://www.iaeng.org/publication/WCE2009/WCE2009\\_pp93-98.pdf](http://www.iaeng.org/publication/WCE2009/WCE2009_pp93-98.pdf)). *Proceedings of the World Congress on Engineering 2009*: 93–98. ISBN 978-988-17-0125-1. .
- [59] Brams, Steven J.; MacLennan, Bruce J.; Täuber, Uwe C.; Manzan, Sebastiano; Mizrach, Bruce; Petkova, Ralitsa; Haas, Markus; Pigorsch, Christian et al. (2009). "Fractal Geometry, A Brief Introduction to". *Encyclopedia of Complexity and Systems Science*. pp. 3700. doi:10.1007/978-0-387-30440-3\_218. ISBN 978-0-387-75888-6.

## References

### Further reading

- Barnsley, Michael F.; and Rising, Hawley; *Fractals Everywhere*. Boston: Academic Press Professional, 1993. ISBN 0-12-079061-0
- Falconer, Kenneth; *Techniques in Fractal Geometry*. John Wiley and Sons, 1997. ISBN 0-471-92287-0
- Jürgens, Hartmut; Peitgen, Heins-Otto; and Saupe, Dietmar; *Chaos and Fractals: New Frontiers of Science*. New York: Springer-Verlag, 1992. ISBN 0-387-97903-4
- Mandelbrot, Benoît B.; *The Fractal Geometry of Nature*. New York: W. H. Freeman and Co., 1982. ISBN 0-7167-1186-9
- Peitgen, Heinz-Otto; and Saupe, Dietmar; eds.; *The Science of Fractal Images*. New York: Springer-Verlag, 1988. ISBN 0-387-96608-0
- Pickover, Clifford A.; ed.; *Chaos and Fractals: A Computer Graphical Journey - A 10 Year Compilation of Advanced Research*. Elsevier, 1998. ISBN 0-444-50002-2
- Jones, Jesse; *Fractals for the Macintosh*, Waite Group Press, Corte Madera, CA, 1993. ISBN 1-878739-46-8.
- Lauwerier, Hans; *Fractals: Endlessly Repeated Geometrical Figures*, Translated by Sophia Gill-Hoffstadt, Princeton University Press, Princeton NJ, 1991. ISBN 0-691-08551-X, cloth. ISBN 0-691-02445-6 paperback. "This book has been written for a wide audience..." Includes sample BASIC programs in an appendix.
- Sprott, Julien Clinton (2003). *Chaos and Time-Series Analysis*. Oxford University Press. ISBN 0-19-850839-5 and ISBN 978-0-19-850839-7.
- Wahl, Bernt; Van Roy, Peter; Larsen, Michael; and Kampman, Eric; *Exploring Fractals on the Macintosh* (<http://www.fractalexplorer.com>), Addison Wesley, 1995. ISBN 0-201-62630-6
- Lesmoir-Gordon, Nigel; "The Colours of Infinity: The Beauty, The Power and the Sense of Fractals." ISBN 1-904555-05-5 (The book comes with a related DVD of the Arthur C. Clarke documentary introduction to the fractal concept and the Mandelbrot set.
- Liu, Huajie; *Fractal Art*, Changsha: Hunan Science and Technology Press, 1997, ISBN 9787535722348.
- Gouyet, Jean-François; *Physics and Fractal Structures* (Foreword by B. Mandelbrot); Masson, 1996. ISBN 2-225-85130-1, and New York: Springer-Verlag, 1996. ISBN 978-0-387-94153-0. Out-of-print. Available in PDF version at. "Physics and Fractal Structures" (<http://www.jfgouyet.fr/fractal/fractauk.html>) (in (French)). Jfgouyet.fr. Retrieved 2010-10-17.
- Bunde, Armin; Havlin, Shlomo (1996). *Fractals and Disordered Systems* ([http://havlin.biu.ac.il/ShlomoHavlinbooks\\_fds.php](http://havlin.biu.ac.il/ShlomoHavlinbooks_fds.php)). Springer.
- Bunde, Armin; Havlin, Shlomo (1995). *Fractals in Science* ([http://havlin.biu.ac.il/ShlomoHavlinbooks\\_f\\_in\\_s.php](http://havlin.biu.ac.il/ShlomoHavlinbooks_f_in_s.php)). Springer.
- ben-Avraham, Daniel; Havlin, Shlomo (2000). *Diffusion and Reactions in Fractals and Disordered Systems* ([http://havlin.biu.ac.il/ShlomoHavlinbooks\\_d\\_r.php](http://havlin.biu.ac.il/ShlomoHavlinbooks_d_r.php)). Cambridge University Press.



## External links

- Fractals ([http://www.dmoz.org/Science/Math/Chaos\\_And\\_Fractals/](http://www.dmoz.org/Science/Math/Chaos_And_Fractals/)) at the Open Directory Project
- Scaling and Fractals (<http://havlin.biu.ac.il/nas1/index.html>) presented by Shlomo Havlin, Bar-Ilan University
- Hunting the Hidden Dimension (<http://www.pbs.org/wgbh/nova/physics/hunting-hidden-dimension.html>), *PBS NOVA*, first aired August 24, 2011
- Benoit Mandelbrot: Fractals and the Art of Roughness ([http://www.ted.com/talks/benoit\\_mandelbrot\\_fractals\\_the\\_art\\_of\\_roughness.html](http://www.ted.com/talks/benoit_mandelbrot_fractals_the_art_of_roughness.html)), TED (conference), February 2010
- Zoom Video in Mandelbox (<http://www.youtube.com/watch?v=7Pf6jZWguCc>)(Exemple of 3D fractal)

# Fibonacci number

In mathematics, the **Fibonacci numbers** or **Fibonacci series** or **Fibonacci sequence** are the numbers in the following integer sequence:<sup>[1][2]</sup>

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ... (sequence A000045 in OEIS)

By definition, the first two numbers in the Fibonacci sequence are 0 and 1, and each subsequent number is the sum of the previous two.

In mathematical terms, the sequence  $F_n$  of Fibonacci numbers is defined by the recurrence relation

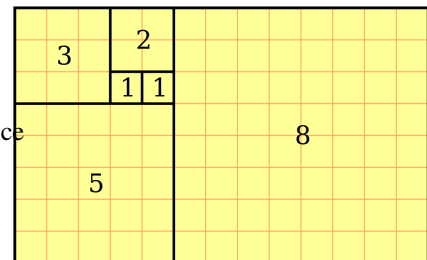
$$F_n = F_{n-1} + F_{n-2},$$

with seed values<sup>[3]</sup>

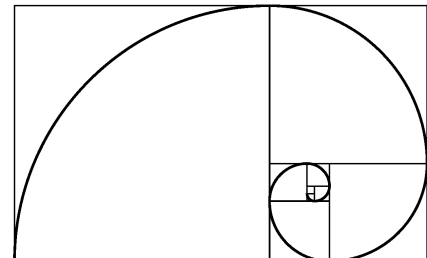
$$F_0 = 0, F_1 = 1.$$

The Fibonacci sequence is named after Leonardo of Pisa, who was known as Fibonacci. Fibonacci's 1202 book *Liber Abaci* introduced the sequence to Western European mathematics,<sup>[4]</sup> although the sequence had been described earlier in Indian mathematics.<sup>[5][6][7]</sup> By modern convention, the sequence begins either with  $F_0 = 0$  or with  $F_1 = 1$ . The *Liber Abaci* began the sequence with  $F_1 = 1$ , without an initial 0.

Fibonacci numbers are closely related to Lucas numbers in that they are a complementary pair of Lucas sequences. They are intimately connected with the golden ratio; for example, the closest rational approximations to the ratio are  $2/1$ ,  $3/2$ ,  $5/3$ ,  $8/5$ , ... . Applications include computer algorithms such as the Fibonacci search technique and the Fibonacci heap data structure, and graphs called Fibonacci cubes used for interconnecting parallel and distributed systems. They also appear in biological settings,<sup>[8]</sup> such as branching in trees, phyllotaxis (the arrangement of leaves on a stem), the fruit sprouts of a pineapple,<sup>[9]</sup> the flowering of artichoke, an uncurling fern and the arrangement of a pine cone.<sup>[10]</sup>



A tiling with squares whose side lengths are successive Fibonacci numbers



An approximation of the golden spiral created by drawing circular arcs connecting the opposite corners of squares in the Fibonacci tiling; this one uses squares of sizes 1, 1, 2, 3, 5, 8, 13, 21, and 34.

## Origins

The Fibonacci sequence appears in Indian mathematics, in connection with Sanskrit prosody.<sup>[6][11]</sup> In the Sanskrit oral tradition, there was much emphasis on how long (L) syllables mix with the short (S), and counting the different patterns of L and S within a given fixed length results in the Fibonacci numbers; the number of patterns that are  $m$  short syllables long is the Fibonacci number  $F_{m+1}$ .<sup>[7]</sup>

Susantha Goonatilake writes that the development of the Fibonacci sequence "is attributed in part to Pingala (200 BC), later being associated with Virahanka (c. 700 AD), Gopāla (c. 1135), and Hemachandra (c. 1150)".<sup>[5]</sup> Parmanand Singh cites Pingala's cryptic formula *misrau cha* ("the two are mixed") and cites scholars who interpret it in context as saying that the cases for  $m$  beats ( $F_{m+1}$ ) is obtained by adding a [S] to  $F_m$  cases and [L] to the  $F_{m-1}$  cases. He dates Pingala before 450 BCE.<sup>[12]</sup>

However, the clearest exposition of the series arises in the work of Virahanka (c. 700 AD), whose own work is lost, but is available in a quotation by Gopala (c. 1135):

Variations of two earlier meters [is the variation]... For example, for [a meter of length] four, variations of meters of two [and] three being mixed, five happens. [works out examples 8, 13, 21]... In this way, the process should be followed in all *mātrā-vrttas* [prosodic combinations].<sup>[13]</sup>

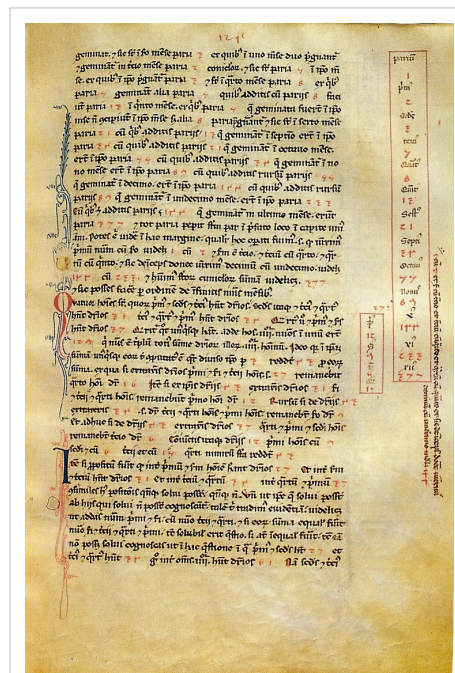
The series is also discussed by Gopala (before 1135 AD) and by the Jain scholar Hemachandra (c. 1150).

In the West, the Fibonacci sequence first appears in the book *Liber Abaci* (1202) by Leonardo of Pisa, known as Fibonacci.<sup>[4]</sup> Fibonacci considers the growth of an idealized (biologically unrealistic) rabbit population, assuming that: a newly born pair of rabbits, one male, one female, are put in a field; rabbits are able to mate at the age of one month so that at the end of its second month a female can produce another pair of rabbits; rabbits never die and a mating pair always produces one new pair (one male, one female) every month from the second month on. The puzzle that Fibonacci posed was: how many pairs will there be in one year?

- At the end of the first month, they mate, but there is still only 1 pair.
- At the end of the second month the female produces a new pair, so now there are 2 pairs of rabbits in the field.
- At the end of the third month, the original female produces a second pair, making 3 pairs in all in the field.
- At the end of the fourth month, the original female has produced yet another new pair, the female born two months ago produces her first pair also, making 5 pairs.

At the end of the  $n$ th month, the number of pairs of rabbits is equal to the number of new pairs (which is the number of pairs in month  $n - 2$ ) plus the number of pairs alive last month ( $n - 1$ ). This is the  $n$ th Fibonacci number.<sup>[14]</sup>

The name "Fibonacci sequence" was first used by the 19th-century number theorist Édouard Lucas.<sup>[15]</sup>



A page of Fibonacci's *Liber Abaci* from the Biblioteca Nazionale di Firenze showing (in box on right) the Fibonacci sequence with the position in the sequence labeled in Roman numerals and the value in Hindu-Arabic numerals.

### List of Fibonacci numbers

The first 21 Fibonacci numbers  $F_n$  for  $n = 0, 1, 2, \dots, 20$  are:<sup>[16]</sup>

$F_0$	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$	$F_9$	$F_{10}$	$F_{11}$	$F_{12}$	$F_{13}$	$F_{14}$	$F_{15}$	$F_{16}$	$F_{17}$	$F_{18}$	$F_{19}$	$F_{20}$
0	1	1	2	3	5	8	13	21	34	55	89	144	233	377	610	987	1597	2584	4181	6765

The sequence can also be extended to negative index  $n$  using the re-arranged recurrence relation

$$F_{n-2} = F_n - F_{n-1},$$

which yields the sequence of "negafibonacci" numbers<sup>[17]</sup> satisfying

$$F_{-n} = (-1)^{n+1} F_n.$$

Thus the bidirectional sequence is

$F_{-8}$	$F_{-7}$	$F_{-6}$	$F_{-5}$	$F_{-4}$	$F_{-3}$	$F_{-2}$	$F_{-1}$	$F_0$	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$
-21	13	-8	5	-3	2	-1	1	0	1	1	2	3	5	8	13	21

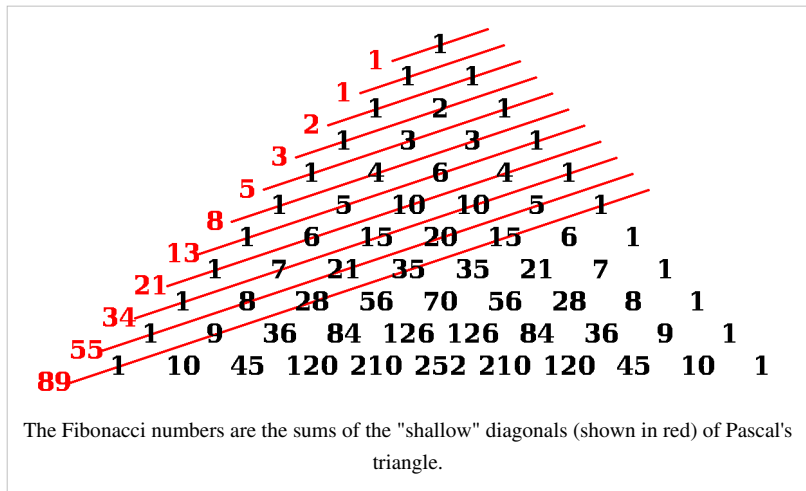
### Occurrences in mathematics

The Fibonacci numbers occur in the sums of "shallow" diagonals in Pascal's triangle (see Binomial coefficient).<sup>[18]</sup>

$$F_n = \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-k-1}{k}.$$

The Fibonacci numbers can be found in different ways in the sequence of binary strings.

- The number of binary strings of length  $n$  without consecutive 1s is the Fibonacci number  $F_{n+2}$ . For example, out of the 16 binary strings of length 4, there are  $F_6 = 8$  without consecutive 1s – they are 0000, 0100, 0010, 0001, 0101, 1000, 1010 and 1001. By symmetry, the number of strings of length  $n$  without consecutive 0s is also  $F_{n+2}$ .
- The number of binary strings of length  $n$  without an odd number of consecutive 1s is the Fibonacci number  $F_{n+1}$ . For example, out of the 16 binary strings of length 4, there are  $F_5 = 5$  without an odd number of consecutive 1s – they are 0000, 0011, 0110, 1100, 1111.
- The number of binary strings of length  $n$  without an even number of consecutive 0s or 1s is  $2F_n$ . For example, out of the 16 binary strings of length 4, there are  $2F_4 = 6$  without an even number of consecutive 0s or 1s – they are 0001, 1000, 1110, 0111, 0101, 1010.



## Relation to the golden ratio

### Closed-form expression

Like every sequence defined by a linear recurrence with constant coefficients, the Fibonacci numbers have a closed-form solution. It has become known as Binet's formula, even though it was already known by Abraham de Moivre:<sup>[19]</sup>

$$F_n = \frac{\varphi^n - \psi^n}{\varphi - \psi} = \frac{\varphi^n - \psi^n}{\sqrt{5}}$$

where

$$\varphi = \frac{1 + \sqrt{5}}{2} \approx 1.61803\,39887\dots$$

is the golden ratio (sequence A001622 in OEIS), and

$$\psi = \frac{1 - \sqrt{5}}{2} = 1 - \varphi = -\frac{1}{\varphi} \approx -0.61803\,39887\dots$$
<sup>[20]</sup>

To see this,<sup>[21]</sup> note that  $\varphi$  and  $\psi$  are both solutions of the equations

$$x^2 = x + 1, \quad x^n = x^{n-1} + x^{n-2},$$

so the powers of  $\varphi$  and  $\psi$  satisfy the Fibonacci recursion. In other words

$$\varphi^n = \varphi^{n-1} + \varphi^{n-2}$$

and

$$\psi^n = \psi^{n-1} + \psi^{n-2}.$$

It follows that for any values  $a$  and  $b$ , the sequence defined by

$$U_n = a\varphi^n + b\psi^n$$

satisfies the same recurrence

$$U_n = a\varphi^{n-1} + b\psi^{n-1} + a\varphi^{n-2} + b\psi^{n-2} = U_{n-1} + U_{n-2}.$$

If  $a$  and  $b$  are chosen so that  $U_0 = 0$  and  $U_1 = 1$  then the resulting sequence  $U_n$  must be the Fibonacci sequence. This is the same as requiring  $a$  and  $b$  satisfy the system of equations:

$$\begin{cases} a + b = 0 \\ \varphi a + \psi b = 1 \end{cases}$$

which has solution

$$a = \frac{1}{\varphi - \psi} = \frac{1}{\sqrt{5}}, \quad b = -a$$

producing the required formula.

### Computation by rounding

Since

$$\frac{|\psi|^n}{\sqrt{5}} < \frac{1}{2}$$

for all  $n \geq 0$ , the number  $F_n$  is the closest integer to

$$\frac{\varphi^n}{\sqrt{5}}.$$

Therefore it can be found by rounding, or in terms of the floor function:

$$F_n = \left\lfloor \frac{\varphi^n}{\sqrt{5}} + \frac{1}{2} \right\rfloor, \quad n \geq 0.$$

Or the nearest integer function:

$$F_n = \left[ \frac{\varphi^n}{\sqrt{5}} \right], \quad n \geq 0.$$

Similarly, if we already know that the number  $F > 1$  is a Fibonacci number, we can determine its index within the sequence by

$$n(F) = \left\lfloor \log_{\varphi} \left( F \cdot \sqrt{5} + \frac{1}{2} \right) \right\rfloor$$

### Limit of consecutive quotients

Johannes Kepler observed that the ratio of consecutive Fibonacci numbers converges. He wrote that "as 5 is to 8 so is 8 to 13, practically, and as 8 is to 13, so is 13 to 21 almost", and concluded that the limit approaches the golden ratio  $\varphi$ .<sup>[22][23]</sup>

$$\lim_{n \rightarrow \infty} \frac{F_{n+1}}{F_n} = \varphi$$

This convergence does not depend on the starting values chosen, excluding 0, 0. For example, the initial values 19 and 31 generate the sequence 19, 31, 50, 81, 131, 212, 343, 555 ... etc. The ratio of consecutive terms in this sequence shows the same convergence towards the golden ratio.

In fact this holds for any sequence which satisfies the Fibonacci recurrence other than a sequence of 0's. This can be derived from Binet's formula.

Another consequence is that the limit of the ratio of two Fibonacci numbers offset by a particular finite deviation in index corresponds to the golden ratio raised by that deviation. Or, in other words:

$$\lim_{n \rightarrow \infty} \frac{F_{n+\alpha}}{F_n} = \varphi^{\alpha}$$

### Decomposition of powers of the golden ratio

Since the golden ratio satisfies the equation

$$\varphi^2 = \varphi + 1,$$

this expression can be used to decompose higher powers  $\varphi^n$  as a linear function of lower powers, which in turn can be decomposed all the way down to a linear combination of  $\varphi$  and 1. The resulting recurrence relationships yield Fibonacci numbers as the linear coefficients:

$$\varphi^n = F_n \varphi + F_{n-1}.$$

This equation can be proved by induction on  $n$ .

This expression is also true for  $n < 1$  if the Fibonacci sequence  $F_n$  is extended to negative integers using the Fibonacci rule  $F_n = F_{n-1} + F_{n-2}$ .

## Matrix form

A 2-dimensional system of linear difference equations that describes the Fibonacci sequence is

$$\begin{pmatrix} F_{k+2} \\ F_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix}$$

$$\vec{F}_{k+1} = A\vec{F}_k$$

The eigenvalues of the matrix  $A$  are  $\varphi = \frac{1}{2}(1+\sqrt{5})$  and  $1-\varphi = \frac{1}{2}(1-\sqrt{5})$ , and the elements of the eigenvectors of  $A$ ,  $\begin{pmatrix} \varphi \\ 1 \end{pmatrix}$  and  $\begin{pmatrix} 1-\varphi \\ 1 \end{pmatrix}$ , are in the ratios  $\varphi$  and  $1-\varphi$ . Using these facts, and the properties of eigenvalues, we can derive a direct formula for the  $n$ th element in the Fibonacci series as an analytic function of  $n$ :

$$F_n = \frac{1}{\sqrt{5}} \cdot \left( \frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \cdot \left( \frac{1-\sqrt{5}}{2} \right)^n$$

The matrix has a determinant of  $-1$ , and thus it is a  $2 \times 2$  unimodular matrix. This property can be understood in terms of the continued fraction representation for the golden ratio:

$$\varphi = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}$$

The Fibonacci numbers occur as the ratio of successive convergents of the continued fraction for  $\varphi$ , and the matrix formed from successive convergents of any continued fraction has a determinant of  $+1$  or  $-1$ .

The matrix representation gives the following closed expression for the Fibonacci numbers:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}.$$

Taking the determinant of both sides of this equation yields Cassini's identity

$$(-1)^n = F_{n+1}F_{n-1} - F_n^2.$$

Additionally, since  $A^n A^m = A^{m+n}$  for any square matrix  $A$ , the following identities can be derived:

$$F_m F_n + F_{m-1} F_{n-1} = F_{m+n-1}$$

$$F_{n+1} F_m + F_n F_{m-1} = F_{m+n}$$

In particular, with  $m = n$ ,

$$F_{2n-1} = F_n^2 + F_{n-1}^2$$

$$\begin{aligned} F_{2n} &= (F_{n-1} + F_{n+1})F_n \\ &= (2F_{n-1} + F_n)F_n \end{aligned}$$

These last two identities provide a way to compute Fibonacci numbers recursively in  $O(\log n)$  arithmetic operations and in time  $O(M(n) \log n)$ , where  $M(n)$  is the time for the multiplication of two numbers of  $n$  digits. This matches the time for computing the  $n$ th Fibonacci number from the closed-form matrix formula, but with fewer redundant steps if one avoids to recompute an already computed Fibonacci number (recursion with memorization).<sup>[24]</sup>

## Recognizing Fibonacci numbers

The question may arise whether a positive integer  $z$  is a Fibonacci number. Since  $F(n)$  is the closest integer to  $\varphi^n / \sqrt{5}$ , the most straightforward, brute-force test is the identity

$$F\left(\left\lfloor \log_{\varphi} \left( z \cdot \sqrt{5} + \frac{1}{2} \right) \right\rfloor\right) = z,$$

which is true if and only if  $z$  is a Fibonacci number. In this formula,  $F(n)$  can be computed rapidly using any of the previously discussed closed-form expressions.

One implication of the above expression is this: if it is known that a number  $z$  is a Fibonacci number, we may determine an  $n$  such that  $F(n) = z$  by the following:

$$\left\lfloor \log_{\varphi} \left( z \cdot \sqrt{5} + \frac{1}{2} \right) \right\rfloor = n$$

Alternatively, a positive integer  $z$  is a Fibonacci number if and only if one of  $5z^2 + 4$  or  $5z^2 - 4$  is a perfect square.<sup>[25]</sup>

A slightly more sophisticated test uses the fact that the convergents of the continued fraction representation of  $\varphi$  are ratios of successive Fibonacci numbers. That is, the inequality

$$\left| \varphi - \frac{p}{q} \right| < \frac{1}{q^2}$$

(with coprime positive integers  $p, q$ ) is true if and only if  $p$  and  $q$  are successive Fibonacci numbers. From this one derives the criterion that  $z$  is a Fibonacci number if and only if the closed interval

$$\left[ \varphi z - \frac{1}{z}, \varphi z + \frac{1}{z} \right]$$

contains a positive integer.<sup>[26]</sup> For  $z \geq 2$ , it is easy to show that this interval contains at most one integer, and in the event that  $z$  is a Fibonacci number, the contained integer is equal to the next successive Fibonacci number after  $z$ . Somewhat remarkably, this result still holds for the case  $z = 1$ , but it must be stated carefully since 1 appears twice in the Fibonacci sequence, and thus has two distinct successors.

## Combinatorial identities

Most identities involving Fibonacci numbers can be proven using combinatorial arguments using the fact that  $F_n$  can be interpreted as the number of sequences of 1s and 2s that sum to  $n - 1$ . This can be taken as the definition of  $F_n$ , with the convention that  $F_0 = 0$ , meaning no sum will add up to  $-1$ , and that  $F_1 = 1$ , meaning the empty sum will "add up" to 0. Here the order of the summand matters. For example,  $1 + 2$  and  $2 + 1$  are considered two different sums.

For example, the recurrence relation

$$F_n = F_{n-1} + F_{n-2},$$

or in words, the  $n$ th Fibonacci number is the sum of the previous two Fibonacci numbers, may be shown by dividing the  $F(n)$  sums of 1s and 2s that add to  $n - 1$  into two non-overlapping groups. One group contains those sums whose first term is 1 and the other those sums whose first term is 2. In the first group the remaining terms add to  $n - 2$ , so it has  $F(n - 1)$  sums, and in the second group the remaining terms add to  $n - 3$ , so there are  $F(n - 2)$  sums. So there are a total of  $F(n - 1) + F(n - 2)$  sums altogether, showing this is equal to  $F(n)$ .

Similarly, it may be shown that the sum of the first Fibonacci numbers up to the  $n$ th is equal to the  $n + 2$ nd Fibonacci number minus 1.<sup>[27]</sup> In symbols:

$$\sum_{i=1}^n F_i = F_{n+2} - 1$$



This is done by dividing the sums adding to  $n+1$  in a different way, this time by the location of the first 2. Specifically, the first group consists of those sums that start with 2, the second group those that start 1+2, the third 1+1+2, and so on, until the last group which consists of the single sum where only 1's are used. The number of sums in the first group is  $F(n)$ ,  $F(n - 1)$  in the second group, and so on, with 1 sum in the last group. So the total number of sums is  $F(n) + F(n - 1) + \dots + F(1)+1$  and therefore this quantity is equal to  $F(n + 2)$

A similar argument, grouping the sums by the position of the first 1 rather than the first 2, gives two more identities:

$$\sum_{i=0}^{n-1} F_{2i+1} = F_{2n}$$

and

$$\sum_{i=1}^n F_{2i} = F_{2n+1} - 1.$$

In words, the sum of the first Fibonacci numbers with odd index up to  $F_{2n-1}$  is the  $(2n)$ th Fibonacci number, and the sum of the first Fibonacci numbers with even index up to  $F_{2n}$  is the  $(2n+1)$ th Fibonacci number minus 1.<sup>[28]</sup>

A different trick may be used to prove

$$\sum_{i=1}^n F_i^2 = F_n F_{n+1},$$

or in words, the sum of the squares of the first Fibonacci numbers up to  $F_n$  is the product of the  $n$ th and  $(n + 1)$ th Fibonacci numbers. In this case note that Fibonacci rectangle of size  $F_n$  by  $F(n + 1)$  can be decomposed into squares of size  $F_n, F_{n-1},$  and so on to  $F_1 = 1$ , from which the identity follows by comparing areas.

### Other identities

There are numerous other identities which can be derived using various methods. Some of the most noteworthy are:<sup>[29]</sup>

$$F_n^2 - F_{n+r}F_{n-r} = (-1)^{n-r} F_r^2 \text{ (Catalan's identity)}$$

$$F_n^2 - F_{n+1}F_{n-1} = (-1)^{n-1} \text{ (Cassini's identity)}$$

$$F_m F_{n+1} - F_{m+1} F_n = (-1)^n F_{m-n} \text{ (d'Ocagne's identity)}$$

$$F_{2n} = F_{n+1}^2 - F_{n-1}^2 = F_n(F_{n+1} + F_{n-1}) = F_n L_n$$

where  $L_n$  is the  $n$ 'th Lucas Number. The last is an identity for doubling  $n$ ; other identities of this type are

$$F_{3n} = 2F_n^3 + 3F_n F_{n+1} F_{n-1} = 5F_n^3 + 3(-1)^n F_n \text{ by Cassini's identity.}$$

$$F_{3n+1} = F_{n+1}^3 + 3F_{n+1} F_n^2 - F_n^3$$

$$F_{3n+2} = F_{n+1}^3 + 3F_{n+1}^2 F_n + F_n^3$$

$$F_{4n} = 4F_n F_{n+1} (F_{n+1}^2 + 2F_n^2) - 3F_n^2 (F_n^2 + 2F_{n+1}^2)$$

These can be found experimentally using lattice reduction, and are useful in setting up the special number field sieve to factorize a Fibonacci number.

More generally,<sup>[29]</sup>

$$F_{kn+c} = \sum_{i=0}^k \binom{k}{i} F_{c-i} F_n^i F_{n+1}^{k-i}.$$

Putting  $k = 2$  in this formula, one gets again the formulas of the end of above section Matrix form.

## Power series

The generating function of the Fibonacci sequence is the power series

$$s(x) = \sum_{k=0}^{\infty} F_k x^k.$$

This series is convergent for  $|x| < \frac{1}{\varphi}$ , and its sum has a simple closed-form:<sup>[30]</sup>

$$s(x) = \frac{x}{1 - x - x^2}$$

This can be proven by using the Fibonacci recurrence to expand each coefficient in the infinite sum:

$$\begin{aligned} s(x) &= \sum_{k=0}^{\infty} F_k x^k \\ &= F_0 + F_1 x + \sum_{k=2}^{\infty} (F_{k-1} + F_{k-2}) x^k \\ &= x + \sum_{k=2}^{\infty} F_{k-1} x^k + \sum_{k=2}^{\infty} F_{k-2} x^k \\ &= x + x \sum_{k=0}^{\infty} F_k x^k + x^2 \sum_{k=0}^{\infty} F_k x^k \\ &= x + x s(x) + x^2 s(x). \end{aligned}$$

Solving the equation  $s(x) = x + x s(x) + x^2 s(x)$  for  $s(x)$  results in the above closed form.

If  $x$  is the inverse of an integer, the closed form of the series becomes

$$\sum_{n=0}^{\infty} \frac{F_n}{k^n} = \frac{k}{k^2 - k - 1}.$$

In particular,

$$\sum_{n=1}^{\infty} \frac{F_n}{10^{(k+1)(n+1)}} = \frac{1}{10^{2k+2} - 10^{k+1} - 1}$$

for all integers  $k \geq 0$ .

Some math puzzle-books present as curious the particular value  $\frac{s(\frac{1}{10})}{10} = \frac{1}{89}$ .<sup>[31]</sup>

## Reciprocal sums

Infinite sums over reciprocal Fibonacci numbers can sometimes be evaluated in terms of theta functions. For example, we can write the sum of every odd-indexed reciprocal Fibonacci number as

$$\sum_{k=0}^{\infty} \frac{1}{F_{2k+1}} = \frac{\sqrt{5}}{4} \vartheta_2^2 \left( 0, \frac{3 - \sqrt{5}}{2} \right),$$

and the sum of squared reciprocal Fibonacci numbers as

$$\sum_{k=1}^{\infty} \frac{1}{F_k^2} = \frac{5}{24} \left( \vartheta_2^4 \left( 0, \frac{3 - \sqrt{5}}{2} \right) - \vartheta_4^4 \left( 0, \frac{3 - \sqrt{5}}{2} \right) + 1 \right).$$

If we add 1 to each Fibonacci number in the first sum, there is also the closed form

$$\sum_{k=0}^{\infty} \frac{1}{1 + F_{2k+1}} = \frac{\sqrt{5}}{2},$$

and there is a nice *nested* sum of squared Fibonacci numbers giving the reciprocal of the golden ratio,

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{\sum_{j=1}^k F_j^2} = \frac{\sqrt{5} - 1}{2}.$$

Results such as these make it plausible that a closed formula for the plain sum of reciprocal Fibonacci numbers could be found, but none is yet known. Despite that, the reciprocal Fibonacci constant

$$\psi = \sum_{k=1}^{\infty} \frac{1}{F_k} = 3.359885666243 \dots$$

has been proved irrational by Richard André-Jeannin.

**Millin series** gives a remarkable identity:<sup>[32]</sup>

$$\sum_{n=0}^{\infty} \frac{1}{F_{2^n}} = \frac{7 - \sqrt{5}}{2}$$

which follows from the closed form for its partial sums as  $N$  tends to infinity:

$$\sum_{n=0}^N \frac{1}{F_{2^n}} = 3 - \frac{F_{2^N - 1}}{F_{2^N}}.$$

## Primes and divisibility

### Divisibility properties

Every 3rd number of the sequence is even and more generally, every  $k$ th number of the sequence is a multiple of  $F_k$ . Thus the Fibonacci sequence is an example of a divisibility sequence. In fact, the Fibonacci sequence satisfies the stronger divisibility property

$$\gcd(F_m, F_n) = F_{\gcd(m, n)}.$$

### Fibonacci primes

A *Fibonacci prime* is a Fibonacci number that is prime. The first few are:

2, 3, 5, 13, 89, 233, 1597, 28657, 514229, ... (sequence A005478 in OEIS).

Fibonacci primes with thousands of digits have been found, but it is not known whether there are infinitely many.<sup>[33]</sup>

$F_{kn}$  is divisible by  $F_n$ , so, apart from  $F_4 = 3$ , any Fibonacci prime must have a prime index. As there are arbitrarily long runs of composite numbers, there are therefore also arbitrarily long runs of composite Fibonacci numbers.

With the exceptions of 1, 8 and 144 ( $F_1 = F_2$ ,  $F_6$  and  $F_{12}$ ) every Fibonacci number has a prime factor that is not a factor of any smaller Fibonacci number (Carmichael's theorem).<sup>[34]</sup>

The only nontrivial square Fibonacci number is 144.<sup>[35]</sup> Attila Pethő proved in 2001 that there is only a finite number of perfect power Fibonacci numbers.<sup>[36]</sup> In 2006, Y. Bugeaud, M. Mignotte, and S. Siksek proved that 8 and 144 are the only such non-trivial perfect powers.<sup>[37]</sup>

No Fibonacci number greater than  $F_6 = 8$  is one greater or one less than a prime number.<sup>[38]</sup>

Any three consecutive Fibonacci numbers, taken two at a time, are relatively prime: that is,

$$\gcd(F_n, F_{n+1}) = \gcd(F_n, F_{n+2}) = 1.$$

More generally,

$$\gcd(F_n, F_m) = F_{\gcd(n, m)}. \quad [39][40]$$

## Prime divisors of Fibonacci numbers

The divisibility of Fibonacci numbers by a prime  $p$  is related to the Legendre symbol  $\left(\frac{p}{5}\right)$  which is evaluated as follows:

$$\left(\frac{p}{5}\right) = \begin{cases} 0 & \text{if } p = 5 \\ 1 & \text{if } p \equiv \pm 1 \pmod{5} \\ -1 & \text{if } p \equiv \pm 2 \pmod{5}. \end{cases}$$

If  $p$  is a prime number then  $F_p \equiv \left(\frac{p}{5}\right) \pmod{p}$  and  $F_{p-\left(\frac{p}{5}\right)} \equiv 0 \pmod{p}$ .<sup>[41][42]</sup>

For example,

$$\begin{aligned} \left(\frac{2}{5}\right) &= -1, & F_3 &= 2, & F_2 &= 1, \\ \left(\frac{3}{5}\right) &= -1, & F_4 &= 3, & F_3 &= 2, \\ \left(\frac{5}{5}\right) &= 0, & F_5 &= 5, \\ \left(\frac{7}{5}\right) &= -1, & F_8 &= 21, & F_7 &= 13, \\ \left(\frac{11}{5}\right) &= +1, & F_{10} &= 55, & F_{11} &= 89. \end{aligned}$$

It is not known whether there exists a prime  $p$  such that  $F_{p-\left(\frac{p}{5}\right)} \equiv 0 \pmod{p^2}$ . Such primes (if there are any) would be called Wall–Sun–Sun primes.

Also, if  $p \neq 5$  is an odd prime number then:<sup>[42]</sup>

$$5F_{(p\pm 1)/2}^2 \equiv \begin{cases} \frac{5\left(\frac{p}{5}\right)\pm 5}{2} \pmod{p} & \text{if } p \equiv 1 \pmod{4} \\ \frac{5\left(\frac{p}{5}\right)\mp 3}{2} \pmod{p} & \text{if } p \equiv 3 \pmod{4}. \end{cases}$$

Examples of all the cases:

$$p = 7 \equiv 3 \pmod{4}, \quad \left(\frac{7}{5}\right) = -1, \quad \frac{5\left(\frac{7}{5}\right) + 3}{2} = -1 \text{ and } \frac{5\left(\frac{7}{5}\right) - 3}{2} = -4.$$

$$F_3 = 2 \text{ and } F_4 = 3.$$

$$5F_3^2 = 20 \equiv -1 \pmod{7} \text{ and } 5F_4^2 = 45 \equiv -4 \pmod{7}$$

$$p = 11 \equiv 3 \pmod{4}, \quad \left(\frac{11}{5}\right) = +1, \quad \frac{5\left(\frac{11}{5}\right) + 3}{2} = 4 \text{ and } \frac{5\left(\frac{11}{5}\right) - 3}{2} = 1.$$

$$F_5 = 5 \text{ and } F_6 = 8.$$

$$5F_5^2 = 125 \equiv 4 \pmod{11} \text{ and } 5F_6^2 = 320 \equiv 1 \pmod{11}$$

$$p = 13 \equiv 1 \pmod{4}, \quad \left(\frac{13}{5}\right) = -1, \quad \frac{5\left(\frac{13}{5}\right) - 5}{2} = -5 \text{ and } \frac{5\left(\frac{13}{5}\right) + 5}{2} = 0.$$

$$F_6 = 8 \text{ and } F_7 = 13.$$

$$5F_6^2 = 320 \equiv -5 \pmod{13} \text{ and } 5F_7^2 = 845 \equiv 0 \pmod{13}$$

$$p = 29 \equiv 1 \pmod{4}, \quad \left(\frac{29}{5}\right) = +1, \quad \frac{5\left(\frac{29}{5}\right) - 5}{2} = 0 \text{ and } \frac{5\left(\frac{29}{5}\right) + 5}{2} = 5.$$

$$F_{14} = 377 \text{ and } F_{15} = 610.$$

$$5F_{14}^2 = 710645 \equiv 0 \pmod{29} \text{ and } 5F_{15}^2 = 1860500 \equiv 5 \pmod{29}$$

For odd  $n$ , all odd prime divisors of  $F_n$  are  $\equiv 1 \pmod{4}$ , implying that all odd divisors of  $F_n$  (as the products of odd prime divisors) are  $\equiv 1 \pmod{4}$ .<sup>[43]</sup>

For example,

$$F_1 = 1, F_3 = 2, F_5 = 5, F_7 = 13, F_9 = 34 = 2 \cdot 17, F_{11} = 89, F_{13} = 233, F_{15} = 610 = 2 \cdot 5 \cdot 61$$

All known factors of Fibonacci numbers  $F(i)$  for all  $i < 50000$  are collected at the relevant repositories.<sup>[44][45]</sup>

## Periodicity modulo $n$

It may be seen that if the members of the Fibonacci sequence are taken mod  $n$ , the resulting sequence must be periodic with period at most  $n^2-1$ . The lengths of the periods for various  $n$  form the so-called Pisano periods (sequence A001175 in OEIS). Determining the Pisano periods in general is an open problem, although for any particular  $n$  it can be solved as an instance of cycle detection.

## Right triangles

Starting with 5, every second Fibonacci number is the length of the hypotenuse of a right triangle with integer sides, or in other words, the largest number in a Pythagorean triple. The length of the longer leg of this triangle is equal to the sum of the three sides of the preceding triangle in this series of triangles, and the shorter leg is equal to the difference between the preceding bypassed Fibonacci number and the shorter leg of the preceding triangle.

The first triangle in this series has sides of length 5, 4, and 3. Skipping 8, the next triangle has sides of length 13, 12 (5 + 4 + 3), and 5 (8 - 3). Skipping 21, the next triangle has sides of length 34, 30 (13 + 12 + 5), and 16 (21 - 5).

This series continues indefinitely. The triangle sides  $a, b, c$  can be calculated directly:

$$\begin{aligned} a_n &= F_{2n-1} \\ b_n &= 2F_n F_{n-1} \\ c_n &= F_n^2 - F_{n-1}^2. \end{aligned}$$

These formulas satisfy  $a_n^2 = b_n^2 + c_n^2$  for all  $n$ , but they only represent triangle sides when  $n > 2$ .

Any four consecutive Fibonacci numbers  $F_n, F_{n+1}, F_{n+2}$  and  $F_{n+3}$  can also be used to generate a Pythagorean triple in a different way<sup>[46]</sup>:

$$a = F_n F_{n+3}; b = 2F_{n+1} F_{n+2}; c = F_{n+1}^2 + F_{n+2}^2; a^2 + b^2 = c^2.$$

Example 1: let the Fibonacci numbers be 1, 2, 3 and 5. Then:

$$\begin{aligned} a &= 1 \times 5 = 5 \\ b &= 2 \times 2 \times 3 = 12 \\ c &= 2^2 + 3^2 = 13 \\ 5^2 + 12^2 &= 13^2. \end{aligned}$$

## Magnitude

Since  $F_n$  is asymptotic to  $\varphi^n / \sqrt{5}$ , the number of digits in  $F_n$  is asymptotic to  $n \log_{10} \varphi \approx 0.2090 n$ . As a consequence, for every integer  $d > 1$  there are either 4 or 5 Fibonacci numbers with  $d$  decimal digits.

More generally, in the base  $b$  representation, the number of digits in  $F_n$  is asymptotic to  $n \log_b \varphi$ .

## Applications

The Fibonacci numbers are important in the computational run-time analysis of Euclid's algorithm to determine the greatest common divisor of two integers: the worst case input for this algorithm is a pair of consecutive Fibonacci numbers.<sup>[47]</sup>

Yuri Matiyasevich was able to show that the Fibonacci numbers can be defined by a Diophantine equation, which led to his original solution of Hilbert's tenth problem.

The Fibonacci numbers are also an example of a complete sequence. This means that every positive integer can be written as a sum of Fibonacci numbers, where any one number is used once at most. Specifically, every positive integer can be written in a unique way as the sum of *one or more* distinct Fibonacci numbers in such a way that the sum does not include any two consecutive Fibonacci numbers. This is known as Zeckendorf's theorem, and a sum of Fibonacci numbers that satisfies these conditions is called a Zeckendorf representation. The Zeckendorf representation of a number can be used to derive its Fibonacci coding.

Fibonacci numbers are used by some pseudorandom number generators.

Fibonacci numbers are used in a polyphase version of the merge sort algorithm in which an unsorted list is divided into two lists whose lengths correspond to sequential Fibonacci numbers – by dividing the list so that the two parts have lengths in the approximate proportion  $\varphi$ . A tape-drive implementation of the polyphase merge sort was described in *The Art of Computer Programming*.

Fibonacci numbers arise in the analysis of the Fibonacci heap data structure.

The Fibonacci cube is an undirected graph with a Fibonacci number of nodes that has been proposed as a network topology for parallel computing.

A one-dimensional optimization method, called the Fibonacci search technique, uses Fibonacci numbers.<sup>[48]</sup>

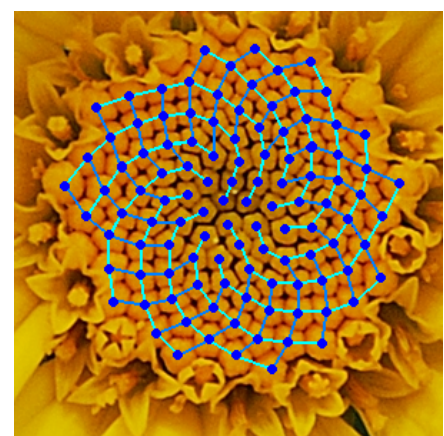
The Fibonacci number series is used for optional lossy compression in the IFF 8SVX audio file format used on Amiga computers. The number series compands the original audio wave similar to logarithmic methods such as  $\mu$ -law.<sup>[49][50]</sup>

Since the conversion factor 1.609344 for miles to kilometers is close to the golden ratio (denoted  $\varphi$ ), the decomposition of distance in miles into a sum of Fibonacci numbers becomes nearly the kilometer sum when the Fibonacci numbers are replaced by their successors. This method amounts to a radix 2 number register in golden ratio base  $\varphi$  being shifted. To convert from kilometers to miles, shift the register down the Fibonacci sequence instead.<sup>[51]</sup>

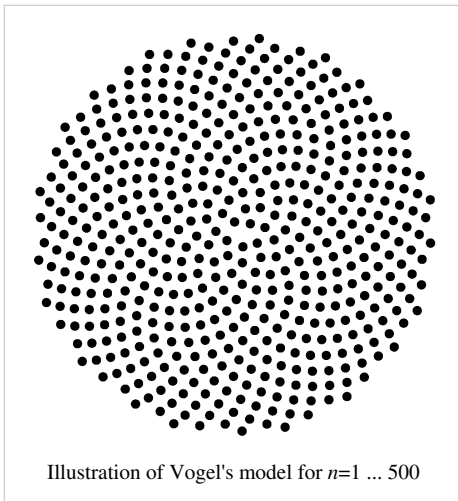
## In nature

Fibonacci sequences appear in biological settings,<sup>[8]</sup> in two consecutive Fibonacci numbers, such as branching in trees, arrangement of leaves on a stem, the fruitlets of a pineapple,<sup>[9]</sup> the flowering of artichoke, an uncurling fern and the arrangement of a pine cone.<sup>[10]</sup> In addition, numerous poorly substantiated claims of Fibonacci numbers or golden sections in nature are found in popular sources, e.g., relating to the breeding of rabbits, the seeds on a sunflower, the spirals of shells, and the curve of waves.<sup>[52]</sup> The Fibonacci numbers are also found in the family tree of honeybees.<sup>[53]</sup>

Przemysław Prusinkiewicz advanced the idea that real instances can in part be understood as the expression of certain algebraic constraints on free groups, specifically as certain Lindenmayer grammars.<sup>[54]</sup>



Yellow Chamomile head showing the arrangement in 21 (blue) and 13 (aqua) spirals. Such arrangements involving consecutive Fibonacci numbers appear in a wide variety of plants.



A model for the pattern of florets in the head of a sunflower was proposed by H. Vogel in 1979.<sup>[55]</sup> This has the form

$$\theta = \frac{2\pi}{\phi^2}n, \quad r = c\sqrt{n}$$

where  $n$  is the index number of the floret and  $c$  is a constant scaling factor; the florets thus lie on Fermat's spiral. The divergence angle, approximately  $137.51^\circ$ , is the golden angle, dividing the circle in the golden ratio. Because this ratio is irrational, no floret has a neighbor at exactly the same angle from the center, so the florets pack efficiently. Because the rational approximations to the golden ratio are of the form  $F(j):F(j+1)$ , the nearest neighbors of floret number  $n$  are those at  $n \pm F(j)$  for some index  $j$  which depends on  $r$ , the distance from the center. It is often said that sunflowers and similar arrangements have

55 spirals in one direction and 89 in the other (or some other pair of adjacent Fibonacci numbers), but this is true only of one range of radii, typically the outermost and thus most conspicuous.<sup>[56]</sup>

### The bee ancestry code

Fibonacci numbers also appear in the description of the reproduction of a population of idealized honeybees, according to the following rules:

- If an egg is laid by an unmated female, it hatches a male or drone bee.
- If, however, an egg was fertilized by a male, it hatches a female.

Thus, a male bee will always have one parent, and a female bee will have two.

If one traces the ancestry of any male bee (1 bee), he has 1 parent (1 bee), 2 grandparents, 3 great-grandparents, 5 great-great-grandparents, and so on. This sequence of numbers of parents is the Fibonacci sequence. The number of ancestors at each level,  $F_n$ , is the number of female ancestors, which is  $F_{n-1}$ , plus the number of male ancestors, which is  $F_{n-2}$ .<sup>[57]</sup> This is under the unrealistic assumption that the ancestors at each level are otherwise unrelated.

### Generalizations

The Fibonacci sequence has been generalized in many ways. These include:

- Generalizing the index to negative integers to produce the Negafibonacci numbers.
- Generalizing the index to real numbers using a modification of Binet's formula.<sup>[29]</sup>
- Starting with other integers. Lucas numbers have  $L_1 = 1, L_2 = 3$ , and  $L_n = L_{n-1} + L_{n-2}$ . Primefree sequences use the Fibonacci recursion with other starting points in order to generate sequences in which all numbers are composite.
- Letting a number be a linear function (other than the sum) of the 2 preceding numbers. The Pell numbers have  $P_n = 2P_{n-1} + P_{n-2}$ .
- Not adding the immediately preceding numbers. The Padovan sequence and Perrin numbers have  $P(n) = P(n-2) + P(n-3)$ .
- Generating the next number by adding 3 numbers (tribonacci numbers), 4 numbers (tetranacci numbers), or more. The resulting sequences are known as *n-Step Fibonacci numbers*.<sup>[58]</sup>
- Adding other objects than integers, for example functions or strings—one essential example is Fibonacci polynomials.



## Notes

- [1] Beck & Geoghegan 2010
- [2] Bona 2011, p. 180
- [3] Lucas 1891, p. 3
- [4] Pisano 2002, pp. 404–5
- [5] Goonatilake, Susantha (1998). *Toward a Global Science* (<http://books.google.com/?id=SI5ip95BbgEC&pg=PA126&dq=Virahanka+Fibonacci>). Indiana University Press. p. 126. ISBN 978-0-253-33388-9. .
- [6] Singh, Parmanand (1985). "The So-called Fibonacci numbers in ancient and medieval India". *Historia Mathematica* **12** (3): 229–44. doi:10.1016/0315-0860(85)90021-7.
- [7] Knuth, Donald (2006). *The Art of Computer Programming* (<http://books.google.com/?id=56LNfE2QGtYC&pg=PA50&dq=rhythms>). **4. Generating All Trees—History of Combinatorial Generation**. Addison–Wesley. p. 50. ISBN 978-0-321-33570-8. . "it was natural to consider the set of all sequences of [L] and [S] that have exactly  $m$  beats. ...there are exactly  $F_{m+1}$  of them. For example the 21 sequences when  $m = 7$  are: [gives list]. In this way Indian prosodists were led to discover the Fibonacci sequence, as we have observed in Section 1.2.8 (from v.1)".
- [8] Douady, S; Couder, Y (1996). "Phyllotaxis as a Dynamical Self Organizing Process" (<http://www.math.ntnu.no/~jarlet/Douady96.pdf>) (PDF). *Journal of Theoretical Biology* **178** (178): 255–74. doi:10.1006/jtbi.1996.0026. .
- [9] Jones, Judy; Wilson, William (2006). "Science". *An Incomplete Education*. Ballantine Books. p. 544. ISBN 978-0-7394-7582-9.
- [10] Brousseau, A (1969). "Fibonacci Statistics in Conifers". *Fibonacci Quarterly* (7): 525–32.
- [11] Knuth, Donald (1968), *The Art of Computer Programming* (<http://books.google.com/?id=MooMkK6ERuYC&pg=PA100&dq=knuth+gopala+fibonacci#v=onepage&>), **1**, Addison Wesley, ISBN 81-7758-754-4, , "Before Fibonacci wrote his work, the sequence  $F_n$  had already been discussed by Indian scholars, who had long been interested in rhythmic patterns... both Gopala (before 1135 AD) and Hemachandra (c. 1150) mentioned the numbers 1,2,3,5,8,13,21 explicitly [see P. Singh Historia Math 12 (1985) 229–44]" p. 100 (3d ed)..."
- [12] Agrawala, VS (1969), Pāṇinikāliṇa Bhāratavarṣa (*Hn.*). *Varanasi-I: TheChowkhamba Vidyabhawan*, "SadgurushiShya writes that Pingala was a younger brother of Pāṇini [Agrawala 1969, lb]. There is an alternative opinion that he was a maternal uncle of Pāṇini [Vinayasagar 1965, Preface, 121. ... Agrawala [1969, 463–76], after a careful investigation in which he considered the views of earlier scholars, has concluded that Pāṇini lived between 480 and 410 BC"
- [13] Velankar, HD (1962). *Ṛttajāṭisamuccaya' of kavi Virahanka*. Jodhpur: Rajasthan Oriental Research Institute. ""For four, variations of meters of two [and] three being mixed, five happens. For five, variations of two earlier – three [and] four, being mixed, eight is obtained. In this way, for six, [variations] of four [and] of five being mixed, thirteen happens. And like that, variations of two earlier meters being mixed, seven morae [is] twenty-one. In this way, the process should be followed in all mātrā-ṛttas".
- [14] Knott, Ron. "Fibonacci's Rabbits" (<http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fibnat.html#Rabbits>). University of Surrey Faculty of Engineering and Physical Sciences. .
- [15] Gardner, Martin (1996). *Mathematical Circus*. The Mathematical Association of America. p. 153. ISBN 0-88385-506-2. "It is ironic that Leonardo, who made valuable contributions to mathematics, is remembered today mainly because a 19th-century French number theorist, Édouard Lucas... attached the name Fibonacci to a number sequence that appears in a trivial problem in Liber abaci".
- [16] Knott, R, "Fib table" (<http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fibtable.html>), *Fibonacci*, UK: Surrey, has the first 300  $F_n$  factored into primes and links to more extensive tables.
- [17] Knuth, Donald (2008-12-11), "Negafibonacci Numbers and the Hyperbolic Plane" ([http://research.allacademic.com/meta/p206842\\_index.html](http://research.allacademic.com/meta/p206842_index.html)), *Annual meeting*, The Fairmont Hotel, San Jose, CA: The Mathematical Association of America, .
- [18] Lucas 1891, p. 7
- [19] Weisstein, Eric W., "Binet's Fibonacci Number Formula (<http://mathworld.wolfram.com/BinetsFibonacciNumberFormula.html>)" from MathWorld.
- [20] Ball 2003, p. 156
- [21] Ball 2003, p. 155–6
- [22] Kepler, Johannes (1966). *A New Year Gift: On Hexagonal Snow*. Oxford University Press. p. 92. ISBN 0-19-858120-3. .
- [23] *Strena seu de Nive Sexangula*, 1611.
- [24] Dijkstra, Edsger W. (1978), *In honour of Fibonacci* (<http://www.cs.utexas.edu/users/EWD/ewd06xx/EWD654.PDF>), .
- [25] Gessel, Ira (October 1972). "Fibonacci is a Square" (<http://www.fq.math.ca/Scanned/10-4/advanced10-4.pdf>) (PDF). *The Fibonacci Quarterly* **10** (4): 417–19. . Retrieved April 11, 2012.
- [26] Möbius, M (1998), "Wie erkennt man eine Fibonacci Zahl?" (in German), *Math. Semesterber* **45**: 243–6.
- [27] Lucas 1891, p. 4
- [28] Vorobiev, Nikolai Nikolaevich; Mircea Martin (2002). "Chapter 1". *Fibonacci Numbers*. Birkhäuser. pp. 5–6. ISBN 3-7643-6135-2.
- [29] Weisstein, Eric W., "Fibonacci Number (<http://mathworld.wolfram.com/FibonacciNumber.html>)" from MathWorld.
- [30] Glaister, P (1995), "Fibonacci power series", *The Mathematical Gazette*: 521.
- [31] Köhler, Günter (February 1985). "Generating functions of Fibonacci-like sequences and decimal expansions of some fractions" (<http://www.fq.math.ca/Scanned/23-1/kohler.pdf>) (PDF). *The Fibonacci Quarterly* **23** (1): 29–35. . Retrieved December 31, 2011.
- [32] Weisstein, Eric W., "Millin Series (<http://mathworld.wolfram.com/MillinSeries.html>)" from MathWorld.
- [33] Weisstein, Eric W., "Fibonacci Prime (<http://mathworld.wolfram.com/FibonacciPrime.html>)" from MathWorld.

- [34] Knott, Ron, *The Fibonacci numbers* (<http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fibtable.html>), UK: Surrey, .
- [35] Cohn, JHE (1964). "Square Fibonacci Numbers etc" (<http://math.la.asu.edu/~checkman/SquareFibonacci.html>). *Fibonacci Quarterly* **2**: 109–13. .
- [36] Pethő, Attila (2001), "Diophantine properties of linear recursive sequences II", *Acta Math. Paedagogicae Nyíregyháziensis* **17**: 81–96.
- [37] Bugeaud, Y; Mignotte, M; Siksek, S (2006), "Classical and modular approaches to exponential Diophantine equations. I. Fibonacci and Lucas perfect powers", *Ann. Math.* **2** (163): 969–1018.
- [38] Honsberger, Ross (1985), "Mathematical Gems III", *AMS Dolciani Mathematical Expositions* (9): 133, ISBN 0-88385-318-3.
- [39] Ribenboim, Paulo (2000), *My Numbers, My Friends*, Springer-Verlag.
- [40] Su, Francis E (2000), "Fibonacci GCD's, please" (<http://www.math.hmc.edu/funfacts/ffiles/20004.5.shtml>), *Mudd Math Fun Facts*, et al, HMC, .
- [41] Ribenboim, Paulo (1996), *The New Book of Prime Number Records*, New York: Springer, p. 64, ISBN 0-387-94457-5.
- [42] Lemmermeyer 2000, pp. 73–4
- [43] Lemmermeyer 2000, p. 73
- [44] *Fibonacci and Lucas factorizations* (<http://mersennus.net/fibonacci/>), Mersennus, collects all known factors of  $F(i)$  with  $i < 10000$ .
- [45] *Factors of Fibonacci and Lucas numbers* (<http://fibonacci.redgolpe.com/>), Red golpe, collects all known factors of  $F(i)$  with  $10000 < i < 50000$ .
- [46] Koshy, Thomas (2007). *Elementary number theory with applications*. Academic Press. p. 581. ISBN 0-12-372487-2.
- [47] Knuth, Donald E (1997). *The Art of Computer Programming. 1: Fundamental Algorithms* (3rd ed.). Addison–Wesley. p. 343. ISBN 0-201-89683-4..
- [48] Avriel, M; Wilde, DJ (1966). "Optimality of the Symmetric Fibonacci Search Technique". *Fibonacci Quarterly* (3): 265–9.
- [49] *Amiga ROM Kernel Reference Manual*, Addison–Wesley, 1991.
- [50] "IFF" ([http://wiki.multimedia.cx/index.php?title=IFF#Fibonacci\\_Delta\\_Compression](http://wiki.multimedia.cx/index.php?title=IFF#Fibonacci_Delta_Compression)), *Multimedia Wiki*, .
- [51] "Zeckendorf representation" ([http://www.encyclopediaofmath.org/index.php/Zeckendorf\\_representation](http://www.encyclopediaofmath.org/index.php/Zeckendorf_representation)), *Encyclopedia of Math.*, .
- [52] Simanek, D. "Fibonacci Flim-Flam" (<http://www.lhup.edu/~dsimanek/pseudo/fibonacci.htm>). LHUP. .
- [53] "Marks for the da Vinci Code: B–" (<http://www.cs4fn.org/maths/bee-davinci.php>). *Maths*. Computer Science For Fun: CS4FN. .
- [54] Prusinkiewicz, Przemyslaw; Hanan, James (1989). *Lindenmayer Systems, Fractals, and Plants (Lecture Notes in Biomathematics)*. Springer-Verlag. ISBN 0-387-97092-4.
- [55] Vogel, H (1979). "A better way to construct the sunflower head". *Mathematical Biosciences* **44** (44): 179–89. doi:10.1016/0025-5564(79)90080-4.
- [56] Prusinkiewicz, Przemyslaw; Lindenmayer, Aristid (1990). *[The Algorithmic Beauty of Plants* (<http://algorithmicbotany.org/papers/#webdocs>)]. Springer-Verlag. pp. 101–7. ISBN 978-0-387-97297-8. .
- [57] *The Fibonacci Numbers and the Ancestry of Bees* (<http://www1.math.american.edu/newstudents/shared/puzzles/fibbee.html>), American, .
- [58] Weisstein, Eric W., "Fibonacci  $n$ -Step Number (<http://mathworld.wolfram.com/Fibonacci-StepNumber.html>)" from MathWorld.

## References

- Ball, Keith M (2003), "8: Fibonacci's Rabbits Revisited", *Strange Curves, Counting Rabbits, and Other Mathematical Explorations*, Princeton, NJ: Princeton University Press, ISBN 0-691-11321-1.
- Beck, Matthias; Geoghegan, Ross (2010), *The Art of Proof: Basic Training for Deeper Mathematics*, New York: Springer.
- Bóna, Miklós (2011), *A Walk Through Combinatorics* (3rd ed.), New Jersey: World Scientific.
- Lemmermeyer, Franz (2000), *Reciprocity Laws*, New York: Springer, ISBN 3-540-66957-4.
- Lucas, Édouard (1891) (in French), *Théorie des nombres*, **1**, Gauthier-Villars.
- Pisano, Leonardo (2002) (hardback), *Fibonacci's Liber Abaci: A Translation into Modern English of the Book of Calculation*, Sources and Studies in the History of Mathematics and Physical Sciences, Sigler, Laurence E, trans, Springer, ISBN 0-387-95419-8, 978-0-387-40737-1 (paperback).

## External links

- Hazewinkel, Michiel, ed. (2001), "Fibonacci numbers" (<http://www.encyclopediaofmath.org/index.php?title=p/f040020>), *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Fibonacci Sequence (<http://www.bbc.co.uk/programmes/b008ct2j>) on *In Our Time* at the BBC. ( listen now ([http://www.bbc.co.uk/iplayer/console/b008ct2j/In\\_Our\\_Time\\_Fibonacci\\_Sequence](http://www.bbc.co.uk/iplayer/console/b008ct2j/In_Our_Time_Fibonacci_Sequence)))
- " Sloane's A000045 : Fibonacci Numbers (<http://oeis.org/A000045>)", *The On-Line Encyclopedia of Integer Sequences*. OEIS Foundation.
- Periods of Fibonacci Sequences Mod m (<http://www.mathpages.com/home/kmath078/kmath078.htm>) at MathPages
- Scientists find clues to the formation of Fibonacci spirals in nature (<http://www.physorg.com/news97227410.html>)
- Implementation to calculate Fibonacci sequence in Lisp (<http://wikinternet.com/wordpress/code/lisp/fibonacci-number/>)

## Golden ratio

The **golden ratio** ( $\varphi$ ) is also called the **golden section** (Latin: *sectio aurea*) or **golden mean**.<sup>[1][2][3]</sup> Other names include **extreme and mean ratio**,<sup>[4]</sup> **medial section**, **divine proportion**, **divine section** (Latin: *sectio divina*), **golden proportion**, **golden cut**,<sup>[5]</sup> **golden number**, and **mean of Phidias**.<sup>[6][7][8]</sup>

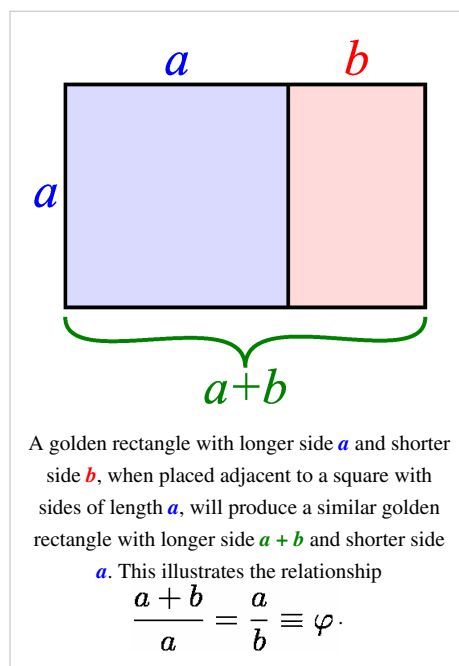
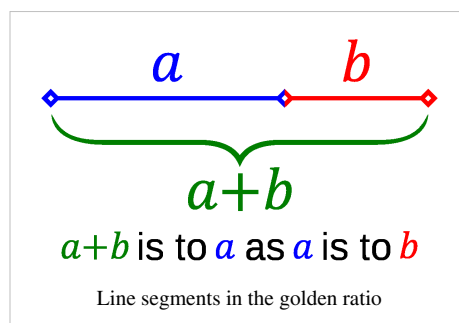
In mathematics and the arts, two quantities are in the golden ratio if the ratio of the sum of the quantities to the larger quantity is equal to the ratio of the larger quantity to the smaller one. The figure on the right illustrates the geometric relationship. Expressed algebraically:

$$\frac{a+b}{a} = \frac{a}{b} \stackrel{\text{def}}{=} \varphi,$$

where the Greek letter phi ( $\varphi$ ) represents the golden ratio. Its value is:

$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.61803\ 39887\ \dots \quad [9]$$

Many 20th century artists and architects have proportioned their works to approximate the golden ratio—especially in the form of the golden rectangle, in which the ratio of the longer side to the shorter is the golden ratio—believing this proportion to be aesthetically pleasing (see Applications and observations below). Mathematicians since Euclid have studied the properties of the golden ratio, including its appearance in the dimensions of a regular pentagon and in a golden rectangle, which can be cut into a square and a smaller rectangle with the same aspect ratio. The golden ratio has also been used to analyze the proportions of natural objects as well as man-made systems such as financial markets, in some cases based on dubious fits to data.<sup>[10]</sup>



### Calculation

<ul style="list-style-type: none"> <li>• List of numbers</li> <li>• Irrational and suspected irrational numbers</li> <li>• <math>\gamma</math></li> <li>• <math>\zeta(3)</math></li> <li>• <math>\sqrt{2}</math></li> <li>• <math>\sqrt{3}</math></li> <li>• <math>\sqrt{5}</math></li> <li>• <math>\varphi</math></li> <li>• <math>\rho</math></li> <li>• <math>\delta_s</math></li> <li>• <math>e</math></li> <li>• <math>\pi</math></li> <li>• <math>\delta</math></li> </ul>	
Binary	1.1001111000110111011...
Decimal	1.6180339887498948482...
Hexadecimal	1.9E3779B97F4A7C15F39...
Continued fraction	$1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}}$
Algebraic form	$\frac{1 + \sqrt{5}}{2}$
Infinite series	$\frac{13}{8} + \sum_{n=0}^{\infty} \frac{(-1)^{(n+1)}(2n+1)!}{(n+2)!n!4^{(2n+3)}}$

Two quantities  $a$  and  $b$  are said to be in the *golden ratio*  $\varphi$  if:

$$\frac{a+b}{a} = \frac{a}{b} = \varphi.$$

One method for finding the value of  $\varphi$  is to start with the left fraction. Through simplifying the fraction and substituting in  $b/a = 1/\varphi$ ,

$$\frac{a+b}{a} = 1 + \frac{b}{a} = 1 + \frac{1}{\varphi},$$

it is shown that

$$1 + \frac{1}{\varphi} = \varphi.$$

Multiplying by  $\varphi$  gives

$$\varphi + 1 = \varphi^2$$

which can be rearranged to

$$\varphi^2 - \varphi - 1 = 0.$$

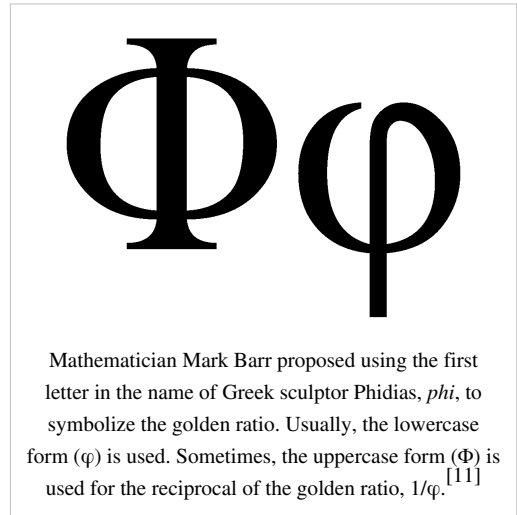
Using the quadratic formula, the only positive solution is

$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.6180339887\dots$$

## History

The golden ratio has fascinated Western intellectuals of diverse interests for at least 2,400 years. According to Mario Livio:

Some of the greatest mathematical minds of all ages, from Pythagoras and Euclid in ancient Greece, through the medieval Italian mathematician Leonardo of Pisa and the Renaissance astronomer Johannes Kepler, to present-day scientific figures such as Oxford physicist Roger Penrose, have spent endless hours over this simple ratio and its properties. But the fascination with the Golden Ratio is not confined just to mathematicians. Biologists, artists, musicians, historians, architects, psychologists, and even mystics have pondered and debated the basis of its ubiquity and appeal. In fact, it is probably fair to say that the Golden Ratio has inspired thinkers of all disciplines like no other number in the history of mathematics.<sup>[12]</sup>



Ancient Greek mathematicians first studied what we now call the golden ratio because of its frequent appearance in geometry. The division of a line into "extreme and mean ratio" (the golden section) is important in the geometry of regular pentagrams and pentagons. The Greeks usually attributed discovery of this concept to Pythagoras or his followers. The regular pentagram, which has a regular pentagon inscribed within it, was the Pythagoreans' symbol.

Euclid's *Elements* (Greek: Στοιχεῖα) provides the first known written definition of what is now called the golden ratio: "A straight line is said to have been *cut in extreme and mean ratio* when, as the whole line is to the greater segment, so is the greater to the less."<sup>[4]</sup> Euclid explains a construction for cutting (sectioning) a line "in extreme and mean ratio", i.e. the golden ratio.<sup>[13]</sup> Throughout the *Elements*, several propositions (theorems in modern terminology) and their proofs employ the golden ratio.<sup>[14]</sup> Some of these propositions show that the golden ratio is an irrational number.

The name "extreme and mean ratio" was the principal term used from the 3rd century BC<sup>[4]</sup> until about the 18th century.

The modern history of the golden ratio starts with Luca Pacioli's *De divina proportione* of 1509, which captured the imagination of artists, architects, scientists, and mystics with the properties, mathematical and otherwise, of the golden ratio.

The first known approximation of the (inverse) golden ratio by a decimal fraction, stated as "about 0.6180340," was written in 1597 by Michael Maestlin of the University of Tübingen in a letter to his former student Johannes Kepler.<sup>[15]</sup>

Since the twentieth century, the golden ratio has been represented by the Greek letter  $\Phi$  or  $\varphi$  (phi, after Phidias, a sculptor who is said to have employed it) or less commonly by  $\tau$  (tau, the first letter of the ancient Greek root  $\tauομή$ —meaning *cut*).<sup>[1][16]</sup>

## Timeline

Timeline according to Priya Hemenway.<sup>[17]</sup>

- Phidias (490–430 BC) made the Parthenon statues that seem to embody the golden ratio.
- Plato (427–347 BC), in his *Timaeus*, describes five possible regular solids (the Platonic solids: the tetrahedron, cube, octahedron, dodecahedron, and icosahedron), some of which are related to the golden ratio.<sup>[18]</sup>
- Euclid (c. 325–c. 265 BC), in his *Elements*, gave the first recorded definition of the golden ratio, which he called, as translated into English, "extreme and mean ratio" (Greek: ἄκρος καὶ μέσος λόγος).<sup>[4]</sup>
- Fibonacci (1170–1250) mentioned the numerical series now named after him in his *Liber Abaci*; the ratio of sequential elements of the Fibonacci sequence approaches the golden ratio asymptotically.
- Luca Pacioli (1445–1517) defines the golden ratio as the "divine proportion" in his *Divina Proportione*.
- Michael Maestlin (1550–1631) publishes the first known approximation of the (inverse) golden ratio as a decimal fraction.
- Johannes Kepler (1571–1630) proves that the golden ratio is the limit of the ratio of consecutive Fibonacci numbers,<sup>[19]</sup> and describes the golden ratio as a "precious jewel": "Geometry has two great treasures: one is the Theorem of Pythagoras, and the other the division of a line into extreme and mean ratio; the first we may compare to a measure of gold, the second we may name a precious jewel." These two treasures are combined in the Kepler triangle.
- Charles Bonnet (1720–1793) points out that in the spiral phyllotaxis of plants going clockwise and counter-clockwise were frequently two successive Fibonacci series.
- Martin Ohm (1792–1872) is believed to be the first to use the term *goldener Schnitt* (golden section) to describe this ratio, in 1835.<sup>[20]</sup>
- Édouard Lucas (1842–1891) gives the numerical sequence now known as the Fibonacci sequence its present name.
- Mark Barr (20th century) suggests the Greek letter phi ( $\varphi$ ), the initial letter of Greek sculptor Phidias's name, as a symbol for the golden ratio.<sup>[21]</sup>
- Roger Penrose (b.1931) discovered a symmetrical pattern that uses the golden ratio in the field of aperiodic tilings, which led to new discoveries about quasicrystals.



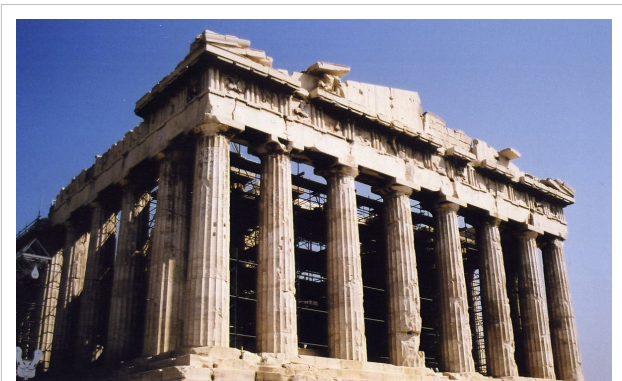
## Applications and observations

### Aesthetics

*De Divina Proportione*, a three-volume work by Luca Pacioli, was published in 1509. Pacioli, a Franciscan friar, was known mostly as a mathematician, but he was also trained and keenly interested in art. *De Divina Proportione* explored the mathematics of the golden ratio. Though it is often said that Pacioli advocated the golden ratio's application to yield pleasing, harmonious proportions, Livio points out that the interpretation has been traced to an error in 1799, and that Pacioli actually advocated the Vitruvian system of rational proportions.<sup>[1]</sup> Pacioli also saw Catholic religious significance in the ratio, which led to his work's title. Containing illustrations of regular solids by Leonardo da Vinci, Pacioli's longtime friend and collaborator, *De Divina Proportione* was a major influence on generations of artists and architects alike.

### Architecture

The Parthenon's façade as well as elements of its façade and elsewhere are said by some to be circumscribed by golden rectangles.<sup>[22]</sup> Other scholars deny that the Greeks had any aesthetic association with golden ratio. For example, Midhat J. Gazalé says, "It was not until Euclid, however, that the golden ratio's mathematical properties were studied. In the *Elements* (308 BC) the Greek mathematician merely regarded that number as an interesting irrational number, in connection with the middle and extreme ratios. Its occurrence in regular pentagons and decagons was duly observed, as well as in the dodecahedron (a regular polyhedron whose twelve faces are regular pentagons). It is indeed



Many of the proportions of the Parthenon are alleged to exhibit the golden ratio.

exemplary that the great Euclid, contrary to generations of mystics who followed, would soberly treat that number for what it is, without attaching to it other than its factual properties."<sup>[23]</sup> And Keith Devlin says, "Certainly, the oft repeated assertion that the Parthenon in Athens is based on the golden ratio is not supported by actual measurements. In fact, the entire story about the Greeks and golden ratio seems to be without foundation. The one thing we know for sure is that Euclid, in his famous textbook *Elements*, written around 300 BC, showed how to calculate its value."<sup>[24]</sup> Near-contemporary sources like Vitruvius exclusively discuss proportions that can be expressed in whole numbers, i.e. commensurate as opposed to irrational proportions.

A 2004 geometrical analysis of earlier research into the Great Mosque of Kairouan reveals a consistent application of the golden ratio throughout the design, according to Boussora and Mazouz.<sup>[25]</sup> They found ratios close to the golden ratio in the overall proportion of the plan and in the dimensioning of the prayer space, the court, and the minaret. The authors note, however, that the areas where ratios close to the golden ratio were found are not part of the original construction, and theorize that these elements were added in a reconstruction.

The Swiss architect Le Corbusier, famous for his contributions to the modern international style, centered his design philosophy on systems of harmony and proportion. Le Corbusier's faith in the mathematical order of the universe was closely bound to the golden ratio and the Fibonacci series, which he described as "rhythms apparent to the eye and clear in their relations with one another. And these rhythms are at the very root of human activities. They resound in man by an organic inevitability, the same fine inevitability which causes the tracing out of the Golden Section by children, old men, savages and the learned."<sup>[26]</sup>



Le Corbusier explicitly used the golden ratio in his Modulor system for the scale of architectural proportion. He saw this system as a continuation of the long tradition of Vitruvius, Leonardo da Vinci's "Vitruvian Man", the work of Leon Battista Alberti, and others who used the proportions of the human body to improve the appearance and function of architecture. In addition to the golden ratio, Le Corbusier based the system on human measurements, Fibonacci numbers, and the double unit. He took suggestion of the golden ratio in human proportions to an extreme: he sectioned his model human body's height at the navel with the two sections in golden ratio, then subdivided those sections in golden ratio at the knees and throat; he used these golden ratio proportions in the Modulor system. Le Corbusier's 1927 Villa Stein in Garches exemplified the Modulor system's application. The villa's rectangular ground plan, elevation, and inner structure closely approximate golden rectangles.<sup>[27]</sup>

Another Swiss architect, Mario Botta, bases many of his designs on geometric figures. Several private houses he designed in Switzerland are composed of squares and circles, cubes and cylinders. In a house he designed in Origgio, the golden ratio is the proportion between the central section and the side sections of the house.<sup>[28]</sup>

In a recent book, author Jason Elliot speculated that the golden ratio was used by the designers of the Naqsh-e Jahan Square and the adjacent Lotfollah mosque.<sup>[29]</sup>

## Painting

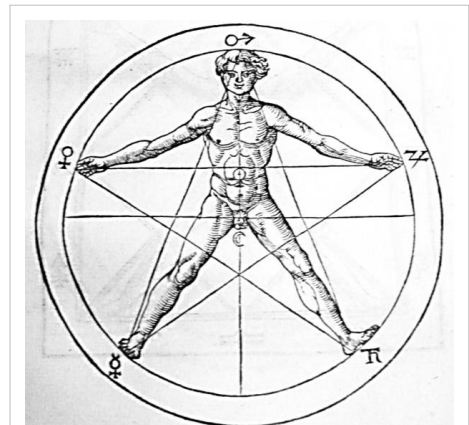
The 16th-century philosopher Heinrich Agrippa drew a man over a pentagram inside a circle, implying a relationship to the golden ratio.<sup>[30]</sup>

Leonardo da Vinci's illustrations of polyhedra in *De divina proportione* (*On the Divine Proportion*) and his views that some bodily proportions exhibit the golden ratio have led some scholars to speculate that he incorporated the golden ratio in his paintings.<sup>[31]</sup> But the suggestion that his *Mona Lisa*, for example, employs golden ratio proportions, is not supported by anything in Leonardo's own writings.<sup>[32]</sup>

Salvador Dalí, influenced by the works of Matila Ghyka,<sup>[33]</sup> explicitly used the golden ratio in his masterpiece, *The Sacrament of the Last Supper*. The dimensions of the canvas are a golden rectangle. A huge dodecahedron, in perspective so that edges appear in golden ratio to one another, is suspended above and behind Jesus and dominates the composition.<sup>[1][34]</sup>

Mondrian has been said to have used the golden section extensively in his geometrical paintings,<sup>[35]</sup> though other experts (including critic Yve-Alain Bois) have disputed this claim.<sup>[1]</sup>

A statistical study on 565 works of art of different great painters, performed in 1999, found that these artists had not used the golden ratio in the size of their canvases. The study concluded that the average ratio of the two sides of the paintings studied is 1.34, with averages for individual artists ranging from 1.04 (Goya) to 1.46 (Bellini).<sup>[36]</sup> On the other hand, Pablo Tosto listed over 350 works by well-known artists, including more than 100 which have canvases with golden rectangle and root-5 proportions, and others with proportions like root-2, 3, 4, and 6.<sup>[37]</sup>



The drawing of a man's body in a pentagram suggests relationships to the golden ratio.<sup>[30]</sup>

## Book design

According to Jan Tschichold,<sup>[39]</sup>

There was a time when deviations from the truly beautiful page proportions 2:3,  $1:\sqrt{3}$ , and the Golden Section were rare. Many books produced between 1550 and 1770 show these proportions exactly, to within half a millimeter.

## Industrial design

Some sources claim that the golden ratio is commonly used in everyday design, for example in the shapes of postcards, playing cards, posters, wide-screen televisions, photographs, and light switch plates.<sup>[40][41][42][43]</sup>

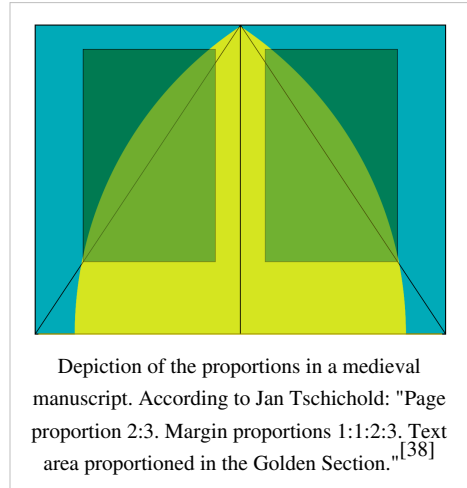
## Music

Ernő Lendvai analyzes Béla Bartók's works as being based on two opposing systems, that of the golden ratio and the acoustic scale,<sup>[44]</sup> though other music scholars reject that analysis.<sup>[1]</sup> In Bartók's *Music for Strings, Percussion and Celesta* the xylophone progression occurs at the intervals 1:2:3:5:8:5:3:2:1.<sup>[45]</sup> French composer Erik Satie used the golden ratio in several of his pieces, including *Sonneries de la Rose+Croix*. The golden ratio is also apparent in the organization of the sections in the music of Debussy's *Reflets dans l'eau* (*Reflections in Water*), from *Images* (1st series, 1905), in which "the sequence of keys is marked out by the intervals 34, 21, 13 and 8, and the main climax sits at the phi position."<sup>[45]</sup>

The musicologist Roy Howat has observed that the formal boundaries of *La Mer* correspond exactly to the golden section.<sup>[46]</sup> Trezise finds the intrinsic evidence "remarkable," but cautions that no written or reported evidence suggests that Debussy consciously sought such proportions.<sup>[47]</sup>

Pearl Drums positions the air vents on its Masters Premium models based on the golden ratio. The company claims that this arrangement improves bass response and has applied for a patent on this innovation.<sup>[48]</sup>

Though Heinz Bohlen proposed the non-octave-repeating 833 cents scale based on combination tones, the tuning features relations based on the golden ratio. As a musical interval the ratio 1.618... is 833.090... cents (Play).<sup>[49]</sup>



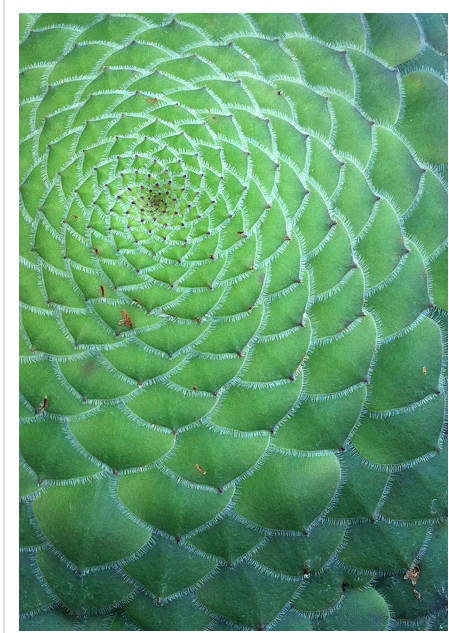
## Nature

Adolf Zeising, whose main interests were mathematics and philosophy, found the golden ratio expressed in the arrangement of branches along the stems of plants and of veins in leaves. He extended his research to the skeletons of animals and the branchings of their veins and nerves, to the proportions of chemical compounds and the geometry of crystals, even to the use of proportion in artistic endeavors. In these phenomena he saw the golden ratio operating as a universal law.<sup>[50]</sup> In connection with his scheme for golden-ratio-based human body proportions, Zeising wrote in 1854 of a universal law "in which is contained the ground-principle of all formative striving for beauty and completeness in the realms of both nature and art, and which permeates, as a paramount spiritual ideal, all structures, forms and proportions, whether cosmic or individual, organic or inorganic, acoustic or optical; which finds its fullest realization, however, in the human form."<sup>[51]</sup>

In 2010, the journal *Science* reported that the golden ratio is present at the atomic scale in the magnetic resonance of spins in cobalt niobate crystals.<sup>[52]</sup>

Several researchers have proposed connections between the golden ratio and human genome DNA.<sup>[53][54][55]</sup>

However, some have argued that many of the apparent manifestations of the golden mean in nature, especially in regard to animal dimensions, are in fact fictitious.<sup>[56]</sup>



A detail of an *Aeonium tabuliforme* in Trädgårdsföreningen, Göteborg

## Optimization

The golden ratio is key to the golden section search.

## Perceptual studies

Studies by psychologists, starting with Fechner, have been devised to test the idea that the golden ratio plays a role in human perception of beauty. While Fechner found a preference for rectangle ratios centered on the golden ratio, later attempts to carefully test such a hypothesis have been, at best, inconclusive.<sup>[1][57]</sup>

## Mathematics

### Golden ratio conjugate

The negative root of the quadratic equation for  $\varphi$  (the "conjugate root") is

$$-\frac{1}{\varphi} = 1 - \varphi = \frac{1 - \sqrt{5}}{2} = -0.61803\ 39887\ \dots$$

The absolute value of this quantity ( $\approx 0.618$ ) corresponds to the length ratio taken in reverse order (shorter segment length over longer segment length,  $b/a$ ), and is sometimes referred to as the *golden ratio conjugate*.<sup>[11]</sup> It is denoted here by the capital Phi ( $\Phi$ ):

$$\Phi = \frac{1}{\varphi} = \frac{1}{1.61803\ 39887\ \dots} = 0.61803\ 39887\ \dots$$

Alternatively,  $\Phi$  can be expressed as

$$\Phi = \varphi - 1 = 1.61803\ 39887\ \dots - 1 = 0.61803\ 39887\ \dots$$

This illustrates the unique property of the golden ratio among positive numbers, that

$$\frac{1}{\varphi} = \varphi - 1,$$

or its inverse:

$$\frac{1}{\Phi} = \Phi + 1.$$

This means  $0.61803\dots:1 = 1:1.61803\dots$

## Short proofs of irrationality

### Contradiction from an expression in lowest terms

Recall that:

the whole is the longer part plus the shorter part;

the whole is to the longer part as the longer part is to the shorter part.

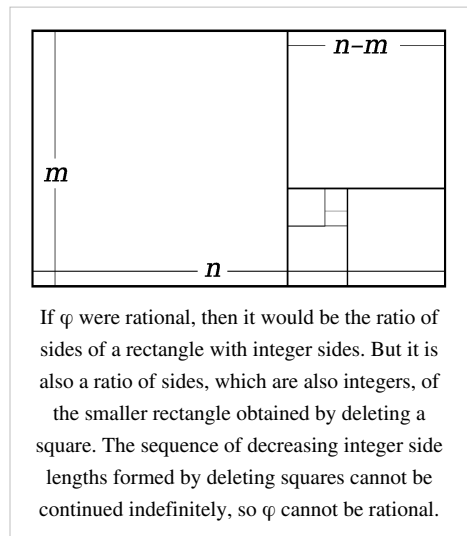
If we call the whole  $n$  and the longer part  $m$ , then the second statement above becomes

$n$  is to  $m$  as  $m$  is to  $n - m$ ,

or, algebraically

$$\frac{n}{m} = \frac{m}{n - m}. \quad (*)$$

To say that  $\varphi$  is rational means that  $\varphi$  is a fraction  $n/m$  where  $n$  and  $m$  are integers. We may take  $n/m$  to be in lowest terms and  $n$  and  $m$  to be positive. But if  $n/m$  is in lowest terms, then the identity labeled (\*) above says  $m/(n - m)$  is in still lower terms. That is a contradiction that follows from the assumption that  $\varphi$  is rational.



If  $\varphi$  were rational, then it would be the ratio of sides of a rectangle with integer sides. But it is also a ratio of sides, which are also integers, of the smaller rectangle obtained by deleting a square. The sequence of decreasing integer side lengths formed by deleting squares cannot be continued indefinitely, so  $\varphi$  cannot be rational.

**Derivation from irrationality of  $\sqrt{5}$**

Another short proof—perhaps more commonly known—of the irrationality of the golden ratio makes use of the closure of rational numbers under addition and multiplication. If  $\frac{1+\sqrt{5}}{2}$  is rational, then  $2\left(\frac{1+\sqrt{5}}{2} - \frac{1}{2}\right) = \sqrt{5}$  is also rational, which is a contradiction if it is already known that the square root of a non-square natural number is irrational.

**Alternative forms**

The formula  $\varphi = 1 + 1/\varphi$  can be expanded recursively to obtain a continued fraction for the golden ratio.<sup>[58]</sup>

$$\varphi = [1; 1, 1, 1, \dots] = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}$$

and its reciprocal:

$$\varphi^{-1} = [0; 1, 1, 1, \dots] = 0 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}$$

The convergents of these continued fractions (1/1, 2/1, 3/2, 5/3, 8/5, 13/8, ...) or (1/1, 1/2, 2/3, 3/5, 5/8, 8/13, ...) are ratios of successive Fibonacci numbers.

The equation  $\varphi^2 = 1 + \varphi$  likewise produces the continued square root, or infinite surd, form:

$$\varphi = \sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{1 + \dots}}}}$$

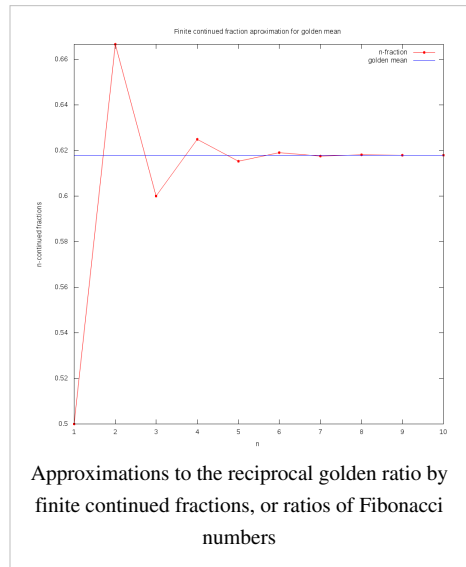
An infinite series can be derived to express phi:<sup>[59]</sup>

$$\varphi = \frac{13}{8} + \sum_{n=0}^{\infty} \frac{(-1)^{(n+1)}(2n+1)!}{(n+2)!n!4^{(2n+3)}}$$

Also:

$$\begin{aligned} \varphi &= 1 + 2 \sin(\pi/10) = 1 + 2 \sin 18^\circ \\ \varphi &= \frac{1}{2} \csc(\pi/10) = \frac{1}{2} \csc 18^\circ \\ \varphi &= 2 \cos(\pi/5) = 2 \cos 36^\circ \\ \varphi &= 2 \sin(3\pi/10) = 2 \sin 54^\circ. \end{aligned}$$

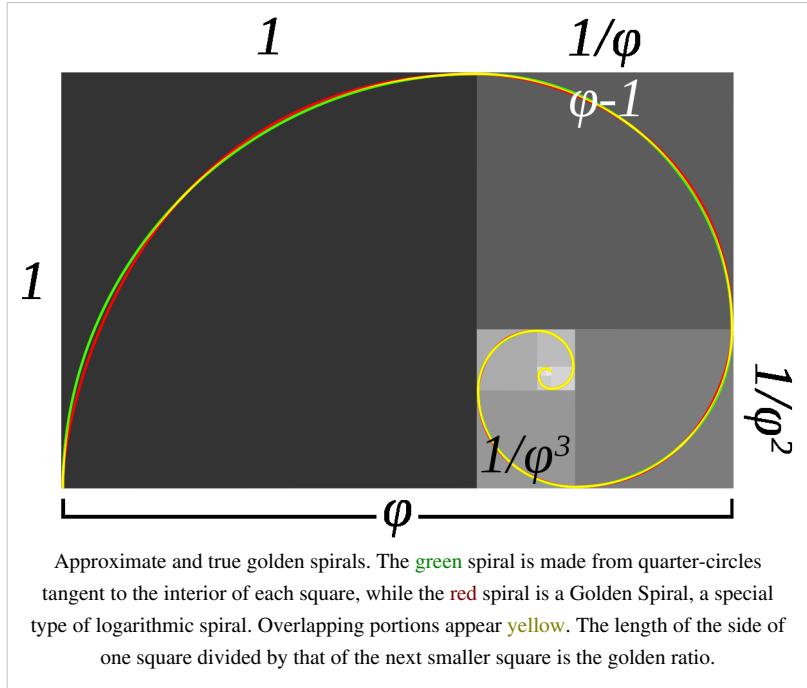
These correspond to the fact that the length of the diagonal of a regular pentagon is  $\varphi$  times the length of its side, and similar relations in a pentagram.



### Geometry

The number  $\varphi$  turns up frequently in geometry, particularly in figures with pentagonal symmetry. The length of a regular pentagon's diagonal is  $\varphi$  times its side. The vertices of a regular icosahedron are those of three mutually orthogonal golden rectangles.

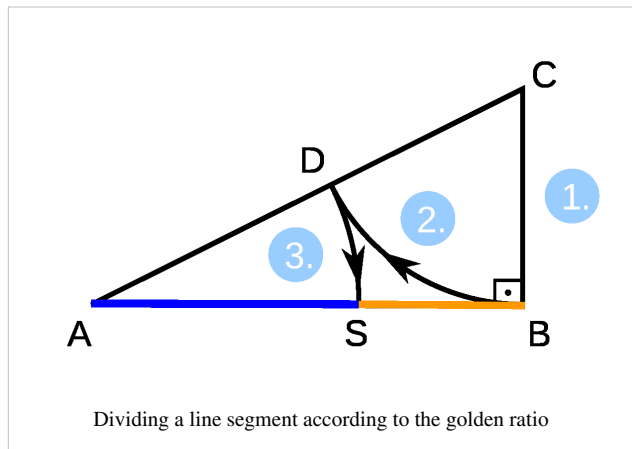
There is no known general algorithm to arrange a given number of nodes evenly on a sphere, for any of several definitions of even distribution (see, for example, *Thomson problem*). However, a useful approximation results from dividing the sphere into parallel bands of equal area and placing one node in each band at longitudes spaced by a golden section of the circle, i.e.  $360^\circ/\varphi \cong 222.5^\circ$ . This method was used to arrange the 1500 mirrors of the student-participatory satellite Starshine-3.<sup>[60]</sup>



### Dividing a line segment

The following algorithm produces a geometric construction that divides a line segment into two line segments where the ratio of the longer to the shorter line segment is the golden ratio:

1. Having a line segment AB, construct a perpendicular BC at point B, with BC half the length of AB. Draw the hypotenuse AC.
2. Draw a circle with center C and radius BC. This circle intersects the hypotenuse AC at point D.
3. Draw a circle with center A and radius AD. This circle intersects the original line segment AB at point S. Point S divides the original segment AB into line segments AS and SB with lengths in the golden ratio.

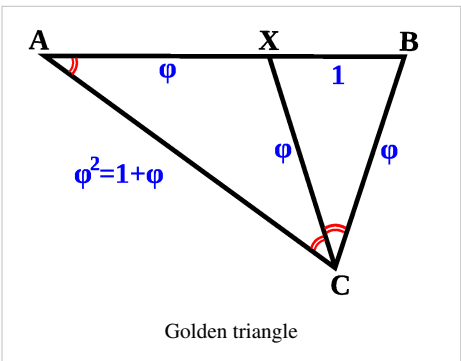


**Golden triangle, pentagon and pentagram**

**Golden triangle**

The golden triangle can be characterized as an isosceles triangle ABC with the property that bisecting the angle C produces a new triangle CXB which is a similar triangle to the original.

If angle BCX =  $\alpha$ , then XCA =  $\alpha$  because of the bisection, and CAB =  $\alpha$  because of the similar triangles; ABC =  $2\alpha$  from the original isosceles symmetry, and BXC =  $2\alpha$  by similarity. The angles in a triangle add up to  $180^\circ$ , so  $5\alpha = 180$ , giving  $\alpha = 36^\circ$ . So the angles of the golden triangle are thus  $36^\circ-72^\circ-72^\circ$ . The angles of the remaining obtuse isosceles triangle AXC (sometimes called the golden gnomon) are  $36^\circ-36^\circ-108^\circ$ .



Suppose XB has length 1, and we call BC length  $\phi$ . Because of the isosceles triangles  $XC=XA$  and  $BC=XC$ , so these are also length  $\phi$ . Length  $AC = AB$ , therefore equals  $\phi + 1$ . But triangle ABC is similar to triangle CXB, so  $AC/BC = BC/BX$ , and so AC also equals  $\phi^2$ . Thus  $\phi^2 = \phi + 1$ , confirming that  $\phi$  is indeed the golden ratio.

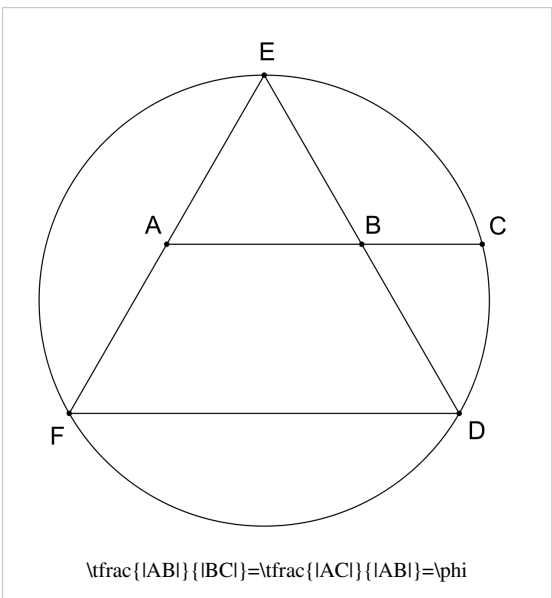
Similarly, the ratio of the area of the larger triangle AXC to the smaller CXB is equal to  $\phi$ , while the inverse ratio is  $\phi - 1$ .

**Pentagon**

In a regular pentagon the ratio between a side and a diagonal is  $\Phi$  (i.e.  $1/\phi$ ), while intersecting diagonals section each other in the golden ratio.<sup>[8]</sup>

**Odom's construction**

George Odom has given a remarkably simple construction for  $\phi$  involving an equilateral triangle: if an equilateral triangle is inscribed in a circle and the line segment joining the midpoints of two sides is produced to intersect the circle in either of two points, then these three points are in golden proportion. This result is a straightforward consequence of the intersecting chords theorem and can be used to construct a regular pentagon, a construction that attracted the attention of the noted Canadian geometer H. S. M. Coxeter who published it in Odom's name as a diagram in the *American Mathematical Monthly* accompanied by the single word "Behold!"<sup>[61]</sup>

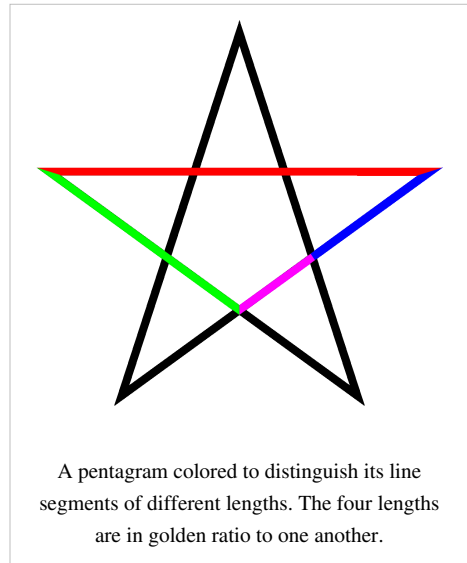




**Pentagram**

The golden ratio plays an important role in the geometry of pentagrams. Each intersection of edges sections other edges in the golden ratio. Also, the ratio of the length of the shorter segment to the segment bounded by the two intersecting edges (a side of the pentagon in the pentagram's center) is  $\varphi$ , as the four-color illustration shows.

The pentagram includes ten isosceles triangles: five acute and five obtuse isosceles triangles. In all of them, the ratio of the longer side to the shorter side is  $\varphi$ . The acute triangles are golden triangles. The obtuse isosceles triangles are golden gnomons.



**Ptolemy's theorem**

The golden ratio properties of a regular pentagon can be confirmed by applying Ptolemy's theorem to the quadrilateral formed by removing one of its vertices. If the quadrilateral's long edge and diagonals are  $b$ , and short edges are  $a$ , then Ptolemy's theorem gives  $b^2 = a^2 + ab$  which yields

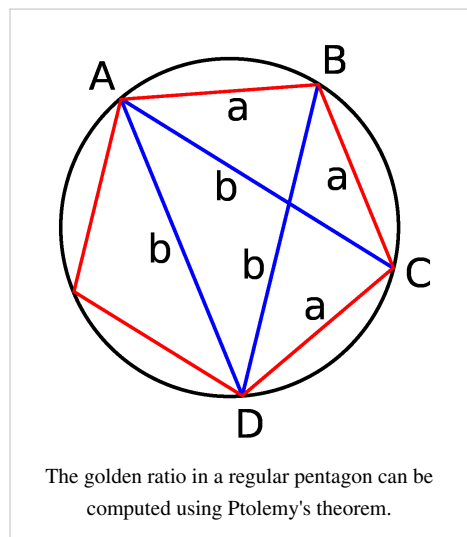
$$\frac{b}{a} = \frac{1 + \sqrt{5}}{2}$$

**Scalenity of triangles**

Consider a triangle with sides of lengths  $a$ ,  $b$ , and  $c$  in decreasing order. Define the "scalenity" of the triangle to be the smaller of the two ratios  $a/b$  and  $b/c$ . The scalenity is always less than  $\varphi$  and can be made as close as desired to  $\varphi$ .<sup>[62]</sup>

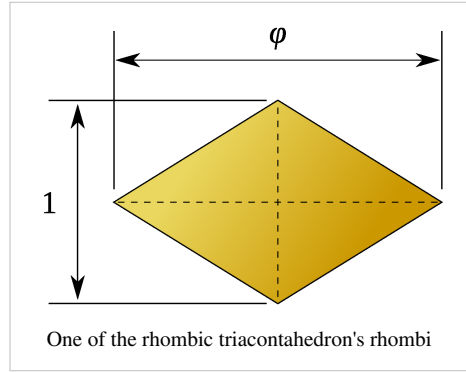
**Triangle whose sides form a geometric progression**

If the side lengths of a triangle form a geometric progression and are in the ratio  $1 : r : r^2$ , where  $r$  is the common ratio, then  $r$  must lie in the range  $\varphi - 1 < r < \varphi$ , which is a consequence of the triangle inequality (the sum of any two sides of a triangle must be strictly bigger than the length of the third side). If  $r = \varphi$  then the shorter two sides are 1 and  $\varphi$  but their sum is  $\varphi^2$ , thus  $r < \varphi$ . A similar calculation shows that  $r > \varphi - 1$ . A triangle whose sides are in the ratio  $1 : \sqrt{\varphi} : \varphi$  is a right triangle (because  $1 + \varphi = \varphi^2$ ) known as a Kepler triangle.<sup>[63]</sup>



**Golden triangle, rhombus, and rhombic triacontahedron**

A golden rhombus is a rhombus whose diagonals are in the golden ratio. The rhombic triacontahedron is a convex polytope that has a very special property: all of its faces are golden rhombi. In the rhombic triacontahedron the dihedral angle between any two adjacent rhombi is  $144^\circ$ , which is twice the isosceles angle of a golden triangle and four times its most acute angle.



One of the rhombic triacontahedron's rhombi

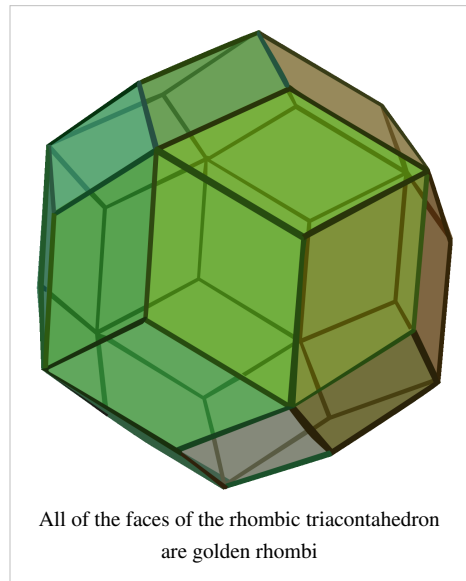
**Relationship to Fibonacci sequence**

The mathematics of the golden ratio and of the Fibonacci sequence are intimately interconnected. The Fibonacci sequence is:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, ...

The closed-form expression (known as Binet's formula, even though it was already known by Abraham de Moivre) for the Fibonacci sequence involves the golden ratio:

$$F(n) = \frac{\phi^n - (1 - \phi)^n}{\sqrt{5}} = \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}}$$



All of the faces of the rhombic triacontahedron are golden rhombi

The golden ratio is the limit of the ratios of successive terms of the Fibonacci sequence (or any Fibonacci-like sequence), as originally shown by Kepler:<sup>[19]</sup>

$$\lim_{n \rightarrow \infty} \frac{F(n+1)}{F(n)} = \phi$$

Therefore, if a Fibonacci number is divided by its immediate predecessor in the sequence, the quotient approximates  $\phi$ ; e.g.,  $987/610 \approx 1.6180327868852$ . These approximations are alternately lower and higher than  $\phi$ , and converge on  $\phi$  as the Fibonacci numbers increase, and:

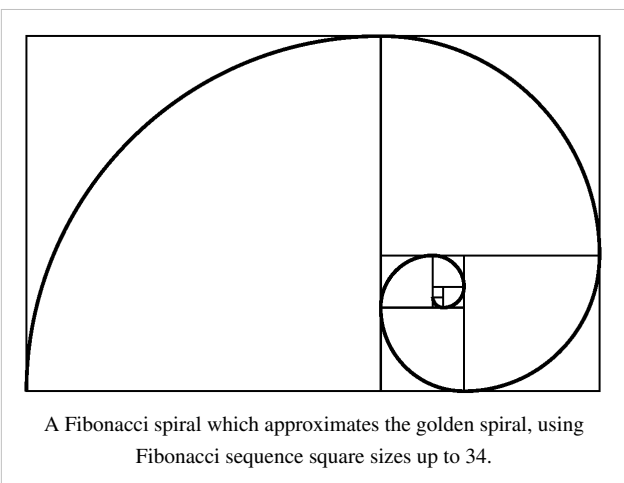
$$\sum_{n=1}^{\infty} |F(n)\phi - F(n+1)| = \phi$$

More generally:

$$\lim_{n \rightarrow \infty} \frac{F(n+a)}{F(n)} = \phi^a$$

where above, the ratios of consecutive terms of the Fibonacci sequence, is a case when  $a = 1$ .

Furthermore, the successive powers of  $\phi$  obey the Fibonacci recurrence:



A Fibonacci spiral which approximates the golden spiral, using Fibonacci sequence square sizes up to 34.

$$\varphi^{n+1} = \varphi^n + \varphi^{n-1}.$$

This identity allows any polynomial in  $\varphi$  to be reduced to a linear expression. For example:

$$\begin{aligned} 3\varphi^3 - 5\varphi^2 + 4 &= 3(\varphi^2 + \varphi) - 5\varphi^2 + 4 \\ &= 3[(\varphi + 1) + \varphi] - 5(\varphi + 1) + 4 \\ &= \varphi + 2 \approx 3.618. \end{aligned}$$

However, this is no special property of  $\varphi$ , because polynomials in any solution  $x$  to a quadratic equation can be reduced in an analogous manner, by applying:

$$x^2 = ax + b$$

for given coefficients  $a, b$  such that  $x$  satisfies the equation. Even more generally, any rational function (with rational coefficients) of the root of an irreducible  $n$ th-degree polynomial over the rationals can be reduced to a polynomial of degree  $n - 1$ . Phrased in terms of field theory, if  $\alpha$  is a root of an irreducible  $n$ th-degree polynomial, then  $\mathbb{Q}(\alpha)$  has degree  $n$  over  $\mathbb{Q}$ , with basis  $\{1, \alpha, \dots, \alpha^{n-1}\}$ .

### Symmetries

The golden ratio and inverse golden ratio  $\varphi_{\pm} = (1 \pm \sqrt{5})/2$  have a set of symmetries that preserve and interrelate them. They are both preserved by the fractional linear transformations  $x, 1/(1 - x), (x - 1)/x, -$  this fact corresponds to the identity and the definition quadratic equation. Further, they are interchanged by the three maps  $1/x, 1 - x, x/(x - 1)$  – they are reciprocals, symmetric about  $1/2$ , and (projectively) symmetric about  $2$ .

More deeply, these maps form a subgroup of the modular group  $\text{PSL}(2, \mathbf{Z})$  isomorphic to the symmetric group on 3 letters,  $S_3$ , corresponding to the stabilizer of the set  $\{0, 1, \infty\}$  of 3 standard points on the projective line, and the symmetries correspond to the quotient map  $S_3 \rightarrow S_2$  – the subgroup  $C_3 < S_3$  consisting of the 3-cycles and the identity  $(\ ) (01\infty)(0\infty 1)$  fixes the two numbers, while the 2-cycles interchange these, thus realizing the map.

### Other properties

The golden ratio has the simplest expression (and slowest convergence) as a continued fraction expansion of any irrational number (see *Alternate forms* above). It is, for that reason, one of the worst cases of Lagrange's approximation theorem and it is an extremal case of the Hurwitz inequality for Diophantine approximations. This may be why angles close to the golden ratio often show up in phyllotaxis (the growth of plants).<sup>[64]</sup>

The defining quadratic polynomial and the conjugate relationship lead to decimal values that have their fractional part in common with  $\varphi$ :

$$\begin{aligned} \varphi^2 &= \varphi + 1 = 2.618\dots \\ \frac{1}{\varphi} &= \varphi - 1 = 0.618\dots \end{aligned}$$

The sequence of powers of  $\varphi$  contains these values 0.618..., 1.0, 1.618..., 2.618...; more generally, any power of  $\varphi$  is equal to the sum of the two immediately preceding powers:

$$\varphi^n = \varphi^{n-1} + \varphi^{n-2} = \varphi \cdot F_n + F_{n-1}.$$

As a result, one can easily decompose any power of  $\varphi$  into a multiple of  $\varphi$  and a constant. The multiple and the constant are always adjacent Fibonacci numbers. This leads to another property of the positive powers of  $\varphi$ :

If  $\lfloor n/2 - 1 \rfloor = m$ , then:

$$\begin{aligned} \varphi^n &= \varphi^{n-1} + \varphi^{n-3} + \dots + \varphi^{n-1-2m} + \varphi^{n-2-2m} \\ \varphi^n - \varphi^{n-1} &= \varphi^{n-2}. \end{aligned}$$

When the golden ratio is used as the base of a numeral system (see Golden ratio base, sometimes dubbed *phinary* or *φ-nary*), every integer has a terminating representation, despite φ being irrational, but every fraction has a non-terminating representation.

The golden ratio is a fundamental unit of the algebraic number field  $\mathbb{Q}(\sqrt{5})$  and is a Pisot–Vijayaraghavan number.<sup>[65]</sup> In the field  $\mathbb{Q}(\sqrt{5})$  we have  $\varphi^n = \frac{L_n + F_n\sqrt{5}}{2}$ , where  $L_n$  is the  $n$ -th Lucas number.

The golden ratio also appears in hyperbolic geometry, as the maximum distance from a point on one side of an ideal triangle to the closer of the other two sides: this distance, the side length of the equilateral triangle formed by the points of tangency of a circle inscribed within the ideal triangle, is  $4 \ln \varphi$ .<sup>[66]</sup>

### Decimal expansion

The golden ratio's decimal expansion can be calculated directly from the expression

$$\varphi = \frac{1 + \sqrt{5}}{2},$$

with  $\sqrt{5} \approx 2.2360679774997896964$ . The square root of 5 can be calculated with the Babylonian method, starting with an initial estimate such as  $x\varphi = 2$  and iterating

$$x_{n+1} = \frac{(x_n + 5/x_n)}{2}$$

for  $n = 1, 2, 3, \dots$ , until the difference between  $x_n$  and  $x_{n-1}$  becomes zero, to the desired number of digits.

The Babylonian algorithm for  $\sqrt{5}$  is equivalent to Newton's method for solving the equation  $x^2 - 5 = 0$ . In its more general form, Newton's method can be applied directly to any algebraic equation, including the equation  $x^2 - x - 1 = 0$  that defines the golden ratio. This gives an iteration that converges to the golden ratio itself,

$$x_{n+1} = \frac{x_n^2 + 1}{2x_n - 1},$$

for an appropriate initial estimate  $x\varphi$  such as  $x\varphi = 1$ . A slightly faster method is to rewrite the equation as  $x - 1 - 1/x = 0$ , in which case the Newton iteration becomes

$$x_{n+1} = \frac{x_n^2 + 2x_n}{x_n^2 + 1}.$$

These iterations all converge quadratically; that is, each step roughly doubles the number of correct digits. The golden ratio is therefore relatively easy to compute with arbitrary precision. The time needed to compute  $n$  digits of the golden ratio is proportional to the time needed to divide two  $n$ -digit numbers. This is considerably faster than known algorithms for the transcendental numbers  $\pi$  and  $e$ .

An easily programmed alternative using only integer arithmetic is to calculate two large consecutive Fibonacci numbers and divide them. The ratio of Fibonacci numbers  $F_{25001}$  and  $F_{25000}$ , each over 5000 digits, yields over 10,000 significant digits of the golden ratio.

The golden ratio  $\varphi$  has been calculated to an accuracy of several millions of decimal digits (sequence A001622 in OEIS). Alexis Irlande performed computations and verification of the first 17,000,000,000 digits.<sup>[67]</sup>

## Pyramids

Both Egyptian pyramids and those mathematical regular square pyramids that resemble them can be analyzed with respect to the golden ratio and other ratios.

### Mathematical pyramids and triangles

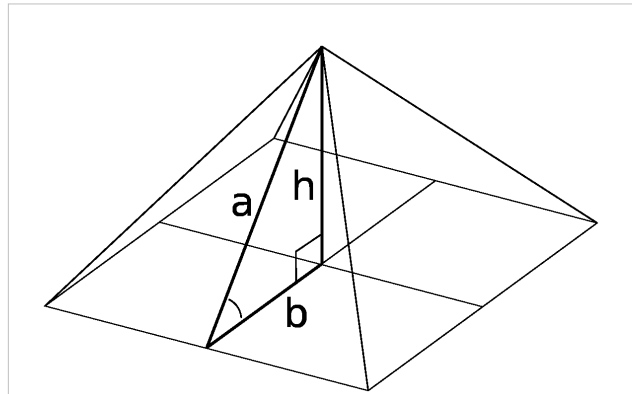
A pyramid in which the apothem (slant height along the bisector of a face) is equal to  $\varphi$  times the semi-base (half the base width) is sometimes called a *golden pyramid*. The isosceles triangle that is the face of such a pyramid can be constructed from the two halves of a diagonally split golden rectangle (of size semi-base by apothem), joining the medium-length edges to make the apothem. The height of this pyramid is  $\sqrt{\varphi}$  times the semi-base (that is, the slope of the face is  $\sqrt{\varphi}$ ); the square of the height is equal to the area of a face,  $\varphi$  times the square of the semi-base.

The medial right triangle of this "golden" pyramid (see diagram), with sides  $1 : \sqrt{\varphi} : \varphi$  is interesting in its own right, demonstrating via the Pythagorean theorem the relationship  $\sqrt{\varphi} = \sqrt{\varphi^2 - 1}$  or  $\varphi = \sqrt{1 + \varphi}$ . This "Kepler triangle"<sup>[68]</sup> is the only right triangle proportion with edge lengths in geometric progression,<sup>[63]</sup> just as the 3–4–5 triangle is the only right triangle proportion with edge lengths in arithmetic progression. The angle with tangent  $\sqrt{\varphi}$  corresponds to the angle that the side of the pyramid makes with respect to the ground, 51.827... degrees (51° 49' 38").<sup>[69]</sup>

A nearly similar pyramid shape, but with rational proportions, is described in the Rhind Mathematical Papyrus (the source of a large part of modern knowledge of ancient Egyptian mathematics), based on the 3:4:5 triangle;<sup>[70]</sup> the face slope corresponding to the angle with tangent 4/3 is 53.13 degrees (53 degrees and 8 minutes).<sup>[71]</sup> The slant height or apothem is 5/3 or 1.666... times the semi-base. The Rhind papyrus has another pyramid problem as well, again with rational slope (expressed as run over rise). Egyptian mathematics did not include the notion of irrational numbers,<sup>[72]</sup> and the rational inverse slope (run/rise, multiplied by a factor of 7 to convert to their conventional units of palms per cubit) was used in the building of pyramids.<sup>[70]</sup>

Another mathematical pyramid with proportions almost identical to the "golden" one is the one with perimeter equal to  $2\pi$  times the height, or  $h:b = 4:\pi$ . This triangle has a face angle of 51.854° (51°51'), very close to the 51.827° of the Kepler triangle. This pyramid relationship corresponds to the coincidental relationship  $\sqrt{\varphi} \approx 4/\pi$ .

Egyptian pyramids very close in proportion to these mathematical pyramids are known.<sup>[71]</sup>



A regular square pyramid is determined by its medial right triangle, whose edges are the pyramid's apothem (a), semi-base (b), and height (h); the face inclination angle is also marked. Mathematical proportions  $b:h:a$  of  $1 : \sqrt{\varphi} : \varphi$  and  $3 : 4 : 5$  and  $1 : 4/\pi : 1.61899$  are of particular interest in relation to Egyptian pyramids.

## Egyptian pyramids

In the mid-nineteenth century, Röber studied various Egyptian pyramids including Khafre, Menkaure and some of the Giza, Sakkara, and Abusir groups, and was interpreted as saying that half the base of the side of the pyramid is the middle mean of the side, forming what other authors identified as the Kepler triangle; many other mathematical theories of the shape of the pyramids have also been explored.<sup>[63]</sup>

One Egyptian pyramid is remarkably close to a "golden pyramid"—the Great Pyramid of Giza (also known as the Pyramid of Cheops or Khufu). Its slope of  $51^\circ 52'$  is extremely close—to the "golden" pyramid inclination of  $51^\circ 50'$  and the  $\pi$ -based pyramid inclination of  $51^\circ 51'$ ; other pyramids at Giza (Chephren,  $52^\circ 20'$ , and Mycerinus,  $50^\circ 47'$ )<sup>[70]</sup> are also quite close. Whether the relationship to the golden ratio in these pyramids is by design or by accident remains open to speculation.<sup>[73]</sup> Several other Egyptian pyramids are very close to the rational 3:4:5 shape.<sup>[71]</sup>

Adding fuel to controversy over the architectural authorship of the Great Pyramid, Eric Temple Bell, mathematician and historian, claimed in 1950 that Egyptian mathematics would not have supported the ability to calculate the slant height of the pyramids, or the ratio to the height, except in the case of the 3:4:5 pyramid, since the 3:4:5 triangle was the only right triangle known to the Egyptians and they did not know the Pythagorean theorem, nor any way to reason about irrationals such as  $\pi$  or  $\varphi$ .<sup>[74]</sup>

Michael Rice<sup>[75]</sup> asserts that principal authorities on the history of Egyptian architecture have argued that the Egyptians were well acquainted with the golden ratio and that it is part of mathematics of the Pyramids, citing Giedon (1957).<sup>[76]</sup> Historians of science have always debated whether the Egyptians had any such knowledge or not, contending rather that its appearance in an Egyptian building is the result of chance.<sup>[77]</sup>

In 1859, the pyramidologist John Taylor claimed that, in the Great Pyramid of Giza, the golden ratio is represented by the ratio of the length of the face (the slope height), inclined at an angle  $\theta$  to the ground, to half the length of the side of the square base, equivalent to the secant of the angle  $\theta$ .<sup>[78]</sup> The above two lengths were about 186.4 and 115.2 meters respectively. The ratio of these lengths is the golden ratio, accurate to more digits than either of the original measurements. Similarly, Howard Vyse, according to Matila Ghyka,<sup>[79]</sup> reported the great pyramid height 148.2 m, and half-base 116.4 m, yielding 1.6189 for the ratio of slant height to half-base, again more accurate than the data variability.

## Disputed observations

Examples of disputed observations of the golden ratio include the following:

- Historian John Man states that the pages of the Gutenberg Bible were "based on the golden section shape". However, according to Man's own measurements, the ratio of height to width was 1.45.<sup>[80]</sup>
- Some specific proportions in the bodies of many animals (including humans<sup>[81][82]</sup>) and parts of the shells of mollusks<sup>[3]</sup> and cephalopods are often claimed to be in the golden ratio. There is a large variation in the real measures of these elements in specific individuals, however, and the proportion in question is often significantly different from the golden ratio.<sup>[81]</sup> The ratio of successive phalangeal bones of the digits and the metacarpal bone has been said to approximate the golden ratio.<sup>[82]</sup> The nautilus shell, the construction of which proceeds in a logarithmic spiral, is often cited, usually with the idea that any logarithmic spiral is related to the golden ratio, but sometimes with the claim that each new chamber is proportioned by the golden ratio relative to the previous one,<sup>[83]</sup> however, measurements of nautilus shells do not support this claim.<sup>[84]</sup>
- The proportions of different plant components (numbers of leaves to branches, diameters of geometrical figures inside flowers) are often claimed to show the golden ratio proportion in several species.<sup>[85]</sup> In practice, there are significant variations between individuals, seasonal variations, and age variations in these species. While the golden ratio may be found in some proportions in some individuals at particular times in their life cycles, there is no consistent ratio in their proportions.

- In investing, some practitioners of technical analysis use the golden ratio to indicate support of a price level, or resistance to price increases, of a stock or commodity; after significant price changes up or down, new support and resistance levels are supposedly found at or near prices related to the starting price via the golden ratio.<sup>[86]</sup> The use of the golden ratio in investing is also related to more complicated patterns described by Fibonacci numbers (e.g. Elliott wave principle and Fibonacci retracement). However, other market analysts have published analyses suggesting that these percentages and patterns are not supported by the data.<sup>[87]</sup>

## References and footnotes

- [1] Livio, Mario (2002). *The Golden Ratio: The Story of Phi, The World's Most Astonishing Number* (<http://books.google.com/books?id=w9dmPwAACAAJ>). New York: Broadway Books. ISBN 0-7679-0815-5. .
- [2] Piotr Sadowski, *The Knight on His Quest: Symbolic Patterns of Transition in Sir Gawain and the Green Knight*, Cranbury NJ: Associated University Presses, 1996
- [3] Richard A Dunlap, *The Golden Ratio and Fibonacci Numbers*, World Scientific Publishing, 1997
- [4] Euclid, *Elements* (<http://aleph0.clarku.edu/~djoyce/java/elements/toc.html>), Book 6, Definition 3.
- [5] Summerson John, *Heavenly Mansions: And Other Essays on Architecture* (New York: W.W. Norton, 1963) p. 37. "And the same applies in architecture, to the rectangles representing these and other ratios (e.g. the 'golden cut'). The sole value of these ratios is that they are intellectually fruitful and suggest the rhythms of modular design."
- [6] Jay Hambidge, *Dynamic Symmetry: The Greek Vase*, New Haven CT: Yale University Press, 1920
- [7] William Lidwell, Kritina Holden, Jill Butler, *Universal Principles of Design: A Cross-Disciplinary Reference*, Gloucester MA: Rockport Publishers, 2003
- [8] Pacioli, Luca. *De divina proportione*, Luca Paganinem de Paganinus de Brescia (Antonio Capella) 1509, Venice.
- [9] The golden ratio can be derived by the quadratic formula, by starting with the first number as 1, then solving for 2nd number  $x$ , where the ratios  $(x + 1)/x = x/1$  or (multiplying by  $x$ ) yields:  $x + 1 = x^2$ , or thus a quadratic equation:  $x^2 - x - 1 = 0$ . Then, by the quadratic formula, for positive  $x = (-b + \sqrt{(b^2 - 4ac)})/(2a)$  with  $a = 1$ ,  $b = -1$ ,  $c = -1$ , the solution for  $x$  is:  $(-(-1) + \sqrt{((-1)^2 - 4 \cdot 1 \cdot (-1))})/(2 \cdot 1)$  or  $(1 + \sqrt{5})/2$ .
- [10] Strogatz, Steven (September 24, 2012). "Me, Myself, and Math: Proportion Control" (<http://opinionator.blogs.nytimes.com/2012/09/24/proportion-control/>). *New York Times*. .
- [11] Weisstein, Eric W., "Golden Ratio Conjugate (<http://mathworld.wolfram.com/GoldenRatioConjugate.html>)" from MathWorld.
- [12] Mario Livio, *The Golden Ratio: The Story of Phi, The World's Most Astonishing Number*, p.6
- [13] Euclid, *Elements* (<http://aleph0.clarku.edu/~djoyce/java/elements/toc.html>), Book 6, Proposition 30.
- [14] Euclid, *Elements* (<http://aleph0.clarku.edu/~djoyce/java/elements/toc.html>), Book 2, Proposition 11; Book 4, Propositions 10–11; Book 13, Propositions 1–6, 8–11, 16–18.
- [15] "The Golden Ratio" ([http://www-history.mcs.st-andrews.ac.uk/HistTopics/Golden\\_ratio.html](http://www-history.mcs.st-andrews.ac.uk/HistTopics/Golden_ratio.html)). *The MacTutor History of Mathematics archive*. . Retrieved 2007-09-18.
- [16] Weisstein, Eric W., "Golden Ratio (<http://mathworld.wolfram.com/GoldenRatio.html>)" from MathWorld.
- [17] Hemenway, Priya (2005). *Divine Proportion: Phi In Art, Nature, and Science*. New York: Sterling. pp. 20–21. ISBN 1-4027-3522-7.
- [18] Plato (360 BC) (Benjamin Jowett trans.). "Timaeus" (<http://classics.mit.edu/Plato/timaeus.html>). The Internet Classics Archive. . Retrieved May 30, 2006.
- [19] James Joseph Tattersall (2005). *Elementary number theory in nine chapters* (<http://books.google.com/?id=QGgLf2oFUyC&pg=PA29&dq=golden-ratio+limit+fibonacci+ratio+kepler&q=golden-ratio+limit+fibonacci+ratio+kepler>) (2nd ed.). Cambridge University Press. p. 28. ISBN 978-0-521-85014-8. .
- [20] Underwood Dudley (1999). *Die Macht der Zahl: Was die Numerologie uns weismachen will* ([http://books.google.com/?id=r6WpMO\\_hREyC&pg=PA245&dq=goldener+Schnitt"+ohm](http://books.google.com/?id=r6WpMO_hREyC&pg=PA245&dq=goldener+Schnitt)). Springer. p. 245. ISBN 3-7643-5978-1. .
- [21] Cook, Theodore Andrea (1979) [1914]. *The Curves of Life* (<http://books.google.com/?id=ea-TStM-07EC&pg=PA420&dq=phi+mark+barr+intitle:The+intitle:Curves+intitle:of+intitle:Life>). New York: Dover Publications. ISBN 0-486-23701-X. .
- [22] Van Mersbergen, Audrey M., "Rhetorical Prototypes in Architecture: Measuring the Acropolis with a Philosophical Polemic", *Communication Quarterly*, Vol. 46 No. 2, 1998, pp 194-213.
- [23] Midhat J. Gazalé , *Gnomon*, Princeton University Press, 1999. ISBN 0-691-00514-1
- [24] Keith J. Devlin *The Math Instinct: Why You're A Mathematical Genius (Along With Lobsters, Birds, Cats, And Dogs)*, p. 108 (<http://books.google.com/books?id=eRD9gYk2r6oC&pg=PA108>). New York: Thunder's Mouth Press, 2005, ISBN 1-56025-672-9
- [25] Boussora, Kenza and Mazouz, Said, *The Use of the Golden Section in the Great Mosque of Kairouan*, Nexus Network Journal, vol. 6 no. 1 (Spring 2004), (<http://www.emis.de/journals/NNJ/BouMaz.html>)
- [26] Le Corbusier, *The Modulor* p. 25, as cited in Padovan, Richard, *Proportion: Science, Philosophy, Architecture* (1999), p. 316, Taylor and Francis, ISBN 0-419-22780-6
- [27] Le Corbusier, *The Modulor*, p. 35, as cited in Padovan, Richard, *Proportion: Science, Philosophy, Architecture* (1999), p. 320. Taylor & Francis. ISBN 0-419-22780-6: "Both the paintings and the architectural designs make use of the golden section".
- [28] Urwin, Simon. *Analysing Architecture* (2003) pp. 154-5, ISBN 0-415-30685-X



- [29] Jason Elliot (2006). *Mirrors of the Unseen: Journeys in Iran* ([http://books.google.com/?id=Gcs4lJux3-4C&pg=PA284&dq=intitle:"Mirrors+of+the+Unseen"+golden-ratio+maidan](http://books.google.com/?id=Gcs4lJux3-4C&pg=PA284&dq=intitle:)). Macmillan. pp. 277, 284. ISBN 978-0-312-30191-0. .
- [30] Piotr Sadowski (1996). *The knight on his quest: symbolic patterns of transition in Sir Gawain and the Green Knight* (<http://books.google.com/books?id=RNfQrs3Ccp4C&pg=PA124>). University of Delaware Press. p. 124. ISBN 978-0-87413-580-0. .
- [31] *Leonardo da Vinci's Polyhedra*, by George W. Hart (<http://www.georgehart.com/virtual-polyhedra/leonardo.html>)
- [32] Livio, Mario. "The golden ratio and aesthetics" (<http://plus.maths.org/issue22/features/golden/>). . Retrieved 2008-03-21.
- [33] Salvador Dalí (2008) (in English) (DVD). *The Dalí Dimension: Decoding the Mind of a Genius* (<http://www.dalidimension.com/eng/index.html>). Media 3.14-TVC-FGSD-IRL-AVRO. .
- [34] Hunt, Carla Herndon and Gilkey, Susan Nicodemus. *Teaching Mathematics in the Block* pp. 44, 47, ISBN 1-883001-51-X
- [35] Bouleau, Charles, *The Painter's Secret Geometry: A Study of Composition in Art* (1963) pp.247-8, Harcourt, Brace & World, ISBN 0-87817-259-9
- [36] Olariu, Agata, *Golden Section and the Art of Painting* Available online (<http://arxiv.org/abs/physics/9908036/>)
- [37] Tosto, Pablo, *La composición áurea en las artes plásticas – El número de oro*, Librería Hachette, 1969, p. 134–144
- [38] Jan Tschichold. *The Form of the Book*, pp.43 Fig 4. "Framework of ideal proportions in a medieval manuscript without multiple columns. Determined by Jan Tschichold 1953. Page proportion 2:3. margin proportions 1:1:2:3, Text area proportioned in the Golden Section. The lower outer corner of the text area is fixed by a diagonal as well."
- [39] Jan Tschichold, *The Form of the Book*, Hartley & Marks (1991), ISBN 0-88179-116-4.
- [40] Jones, Ronald (1971). "The golden section: A most remarkable measure". *The Structurist* **11**: 44–52. "Who would suspect, for example, that the switch plate for single light switches are standardized in terms of a Golden Rectangle?"
- [41] Art Johnson (1999). *Famous problems and their mathematicians* (<http://books.google.com/?id=STKX4qadFTkC&pg=PA45&dq=switch+golden+ratio#v=onepage&q=switch+golden+ratio&f=false>). Libraries Unlimited. p. 45. ISBN 978-1-56308-446-1. . "The Golden Ratio is a standard feature of many modern designs, from postcards and credit cards to posters and light-switch plates."
- [42] Alexey Stakhov, Scott Olsen, Scott Anthony Olsen (2009). *The mathematics of harmony: from Euclid to contemporary mathematics and computer science* (<http://books.google.com/?id=K6fac9RxxREC&pg=PA21&dq=credit+card+golden+ratio+rectangle#v=onepage&q=credit+card+golden+ratio+rectangle&f=false>). World Scientific. p. 21. ISBN 978-981-277-582-5. . "A credit card has a form of the golden rectangle."
- [43] Simon Cox (2004). *Cracking the Da Vinci code: the unauthorized guide to the facts behind Dan Brown's bestselling novel* (<http://books.google.com/?id=TbjwhwLCEeAC&q=golden+ratio+postcard&dq=golden+ratio+postcard>). Barnes & Noble Books. ISBN 978-0-7607-5931-8. . "The Golden Ratio also crops up in some very unlikely places: widescreen televisions, postcards, credit cards and photographs all commonly conform to its proportions."
- [44] Lendvai, Ernő (1971). *Béla Bartók: An Analysis of His Music*. London: Kahn and Averill.
- [45] Smith, Peter F. *The Dynamics of Delight: Architecture and Aesthetics* ([http://books.google.com/books?id=ZgftUKoMnpkC&pg=PA83&dq=bartok+intitle:The+intitle:Dynamics+intitle:of+intitle:Delight+intitle:Architecture+intitle:and+intitle:Aesthetics&as\\_brr=0&ei=WkkSR5L6OI--ogLpmoyzBg&sig=Ijw4YifrLhkcdQSMVAjSL5g4zVk](http://books.google.com/books?id=ZgftUKoMnpkC&pg=PA83&dq=bartok+intitle:The+intitle:Dynamics+intitle:of+intitle:Delight+intitle:Architecture+intitle:and+intitle:Aesthetics&as_brr=0&ei=WkkSR5L6OI--ogLpmoyzBg&sig=Ijw4YifrLhkcdQSMVAjSL5g4zVk)) (New York: Routledge, 2003) pp 83, ISBN 0-415-30010-X
- [46] Roy Howat (1983). *Debussy in Proportion: A Musical Analysis* (<http://books.google.com/?id=4bwKykNp24wC&pg=PA169&dq=intitle:Debussy+intitle:in+intitle:Proportion+golden+la-mer>). Cambridge University Press. ISBN 0-521-31145-4. .
- [47] Simon Trezise (1994). *Debussy: La Mer* ([http://books.google.com/?id=THDInge\\_UzcC&pg=PA53&dq=inauthor:Trezise+golden+evidence](http://books.google.com/?id=THDInge_UzcC&pg=PA53&dq=inauthor:Trezise+golden+evidence)). Cambridge University Press. p. 53. ISBN 0-521-44656-2. .
- [48] "Pearl Masters Premium" (<http://www.pearldrums.com/premium-birch.asp>). Pearl Corporation. . Retrieved December 2, 2007.
- [49] "An 833 Cents Scale: An experiment on harmony" (<http://www.huygens-fokker.org/bpsite/833cent.html>), *Huygens-Fokker.org*. Accessed December 1, 2012.
- [50] Richard Padovan (1999). *Proportion* ([http://books.google.com/?id=Vk\\_CQULdAssC&pg=PA306&dq=contained+the+ground-principle+of+all+formative+striving](http://books.google.com/?id=Vk_CQULdAssC&pg=PA306&dq=contained+the+ground-principle+of+all+formative+striving)). Taylor & Francis. pp. 305–306. ISBN 978-0-419-22780-9. .
- [51] Zeising, Adolf, *Neue Lehre van den Proportionen des menschlichen Körpers*, Leipzig, 1854, preface.
- [52] "Golden ratio discovered in a quantum world" ([http://www.eurekalert.org/pub\\_releases/2010-01/haog-grd010510.php](http://www.eurekalert.org/pub_releases/2010-01/haog-grd010510.php)). Eurekalert.org. 2010-01-07. . Retrieved 2011-10-31.
- [53] J.C. Perez (1991), "Chaos DNA and Neuro-computers: A Golden Link" (<http://golden-ratio-in-dna.blogspot.com/2008/01/1991-first-publication-related-to.html>), in *Speculations in Science and Technology* vol. 14 no. 4, ISSN 0155-7785.
- [54] Yamagishi, Michel E.B., and Shimabukuro, Alex I. (2007), "Nucleotide Frequencies in Human Genome and Fibonacci Numbers" (<http://www.springerlink.com/content/p140352473151957/?p=d5b18a2dfce949858e2062449e9ccfad&pi=0>), in *Bulletin of Mathematical Biology*, ISSN 0092-8240 (print), ISSN 1522-9602 (online). PDF full text (<http://www.springerlink.com/content/p140352473151957/fulltext.pdf>)
- [55] Perez, J.-C. (September 2010). "Codon populations in single-stranded whole human genome DNA are fractal and fine-tuned by the Golden Ratio 1.618". *Interdisciplinary Sciences: Computational Life Science* **2** (3): 228–240. doi:10.1007/s12539-010-0022-0. PMID 20658335.
- [56] Pommersheim, James E., Tim K. Marks, and Erica L. Flapan, eds. 2010. *Number Theory: A lively Introduction with Proofes, Applications, and Stories*. John Wiley and Sons: 82.
- [57] The golden ratio and aesthetics (<http://plus.maths.org/issue22/features/golden/>), by Mario Livio.
- [58] Max. Hailperin, Barbara K. Kaiser, and Karl W. Knight (1998). *Concrete Abstractions: An Introduction to Computer Science Using Scheme* (<http://books.google.com/?id=yYvRVueWIZ8C&pg=PA63&dq=continued-fraction+substitute+golden-ratio>). Brooks/Cole Pub. Co.

- ISBN 0-534-95211-9. .
- [59] Brian Roselle, "Golden Mean Series" (<http://sites.google.com/site/goldenmeanseries/>)
- [60] "A Disco Ball in Space" ([http://science.nasa.gov/science-news/science-at-nasa/2001/ast09oct\\_1/](http://science.nasa.gov/science-news/science-at-nasa/2001/ast09oct_1/)). NASA. 2001-10-09. . Retrieved 2007-04-16.
- [61] Chris and Penny. "Quandaries and Queries" (<http://mathcentral.uregina.ca/qq/database/qq.09.02/mary1.html>). Math Central. . Retrieved 23 October 2011.
- [62] *American Mathematical Monthly*, pp. 49-50, 1954.
- [63] Roger Herz-Fischler (2000). *The Shape of the Great Pyramid* (<http://books.google.com/?id=066T3YLuhA0C&pg=PA81&dq=kepler-triangle+geometric>). Wilfrid Laurier University Press. ISBN 0-88920-324-5. .
- [64] Fibonacci Numbers and Nature - Part 2 : Why is the Golden section the "best" arrangement? (<http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fibnat2.html>), from Dr. Ron Knott's (<http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/>) Fibonacci Numbers and the Golden Section (<http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/>), retrieved 2012-11-29.
- [65] Weisstein, Eric W., "Pisot Number" (<http://mathworld.wolfram.com/PisotNumber.html>) from MathWorld.
- [66] Horocycles exinscrits : une propriété hyperbolique remarquable (<http://www.cabri.net/abracadabri/GeoNonE/GeoHyper/KBModele/Biss3KB.html>), cabri.net, retrieved 2009-07-21.
- [67] *The golden number to 17 000 000 000 digits* (<http://www.matematicas.unal.edu.co/airlande/phi.html.en>). Universidad Nacional de Colombia. 2008. .
- [68] Radio, Astraea Web (2006). *The Best of Astraea: 17 Articles on Science, History and Philosophy* (<http://books.google.com/?id=LDTpvbXLxgQC&pg=PA93&dq=kepler-triangle>). Astrea Web Radio. ISBN 1-4259-7040-0. .
- [69] Midhat Gazale, *Gnomon: From Pharaohs to Fractals*, Princeton Univ. Press, 1999
- [70] Eli Maor, *Trigonometric Delights*, Princeton Univ. Press, 2000
- [71] "The Great Pyramid, The Great Discovery, and The Great Coincidence" (<http://www.petrospec-technologies.com/Herkommer/pyramid/pyramid.htm>). . Retrieved 2007-11-25.
- [72] Lancelot Hogben, *Mathematics for the Million*, London: Allen & Unwin, 1942, p. 63., as cited by Dick Teresi, *Lost Discoveries: The Ancient Roots of Modern Science—from the Babylonians to the Maya*, New York: Simon & Schuster, 2003, p.56
- [73] Burton, David M. (1999). *The history of mathematics: an introduction* (<http://books.google.com/books?id=GKtFAAAAYAAJ>) (4 ed.). WCB McGraw-Hill. p. 56. ISBN 0-07-009468-3. .
- [74] Eric Temple Bell, *The Development of Mathematics*, New York: Dover, 1940, p.40
- [75] Rice, Michael, *Egypt's Legacy: The Archetypes of Western Civilisation, 3000 to 30 B.C* pp. 24 Routledge, 2003, ISBN 0-415-26876-1
- [76] S. Giedon, 1957, The Beginnings of Architecture, The A.W. Mellon Lectures in the Fine Arts, 457, as cited in Rice, Michael, *Egypt's Legacy: The Archetypes of Western Civilisation, 3000 to 30 B.C* pp.24 Routledge, 2003
- [77] Markowsky, George (January 1992). "Misconceptions about the Golden Ratio" (<http://www.umcs.maine.edu/~markov/GoldenRatio.pdf>) (PDF). *College Mathematics Journal* (Mathematical Association of America) **23** (1): 2–19. doi:10.2307/2686193. JSTOR 2686193. .
- [78] Taylor, *The Great Pyramid: Why Was It Built and Who Built It?*, 1859
- [79] Matila Ghyka *The Geometry of Art and Life*, New York: Dover, 1977
- [80] Man, John, *Gutenberg: How One Man Remade the World with Word* (2002) pp. 166–167, Wiley, ISBN 0-471-21823-5. "The half-folio page (30.7 × 44.5 cm) was made up of two rectangles—the whole page and its text area—based on the so called 'golden section', which specifies a crucial relationship between short and long sides, and produces an irrational number, as pi is, but is a ratio of about 5:8."
- [81] Pheasant, Stephen (1998). *Bodyspace*. London: Taylor & Francis. ISBN 0-7484-0067-2.
- [82] van Laack, Walter (2001). *A Better History Of Our World: Volume 1 The Universe*. Aachen: van Laach GmbH.
- [83] Ivan Moscovich, *Ivan Moscovich Mastermind Collection: The Hinged Square & Other Puzzles*, New York: Sterling, 2004
- [84] Peterson, Ivars. "Sea shell spirals" ([http://www.sciencenews.org/view/generic/id/6030/title/Sea\\_Shell\\_Spirals](http://www.sciencenews.org/view/generic/id/6030/title/Sea_Shell_Spirals)). *Science News*. .
- [85] Derek Thomas, *Architecture and the Urban Environment: A Vision for the New Age*, Oxford: Elsevier, 2002
- [86] For instance, Osler writes that "38.2 percent and 61.8 percent retracements of recent rises or declines are common," in Osler, Carol (2000). "Support for Resistance: Technical Analysis and Intraday Exchange Rates" (<http://ftp.ny.frb.org/research/epr/00v06n2/0007osle.pdf>) (PDF). *Federal Reserve Bank of New York Economic Policy Review* **6** (2): 53–68. .
- [87] Roy Batchelor and Richard Ramyar, " Magic numbers in the Dow (<http://www.webcitation.org/5reh6NujR>)," 25th International Symposium on Forecasting, 2005, p. 13, 31. " Not since the 'big is beautiful' days have giants looked better (<http://www.telegraph.co.uk/finance/2947908/Not-since-the-big-is-beautiful-days-have-giants-looked-better.html>)", Tom Stevenson, The Daily Telegraph, Apr. 10, 2006, and "Technical failure", The Economist, Sep. 23, 2006, are both popular-press accounts of Batchelor and Ramyar's research.

## Further reading

- Doczi, György (2005) [1981]. *The Power of Limits: Proportional Harmonies in Nature, Art, and Architecture*. Boston: Shambhala Publications. ISBN 1-59030-259-1.
- Huntley, H. E. (1970). *The Divine Proportion: A Study in Mathematical Beauty*. New York: Dover Publications. ISBN 0-486-22254-3.
- Livio, Mario (2002) [2002]. *The Golden Ratio: The Story of PHI, the World's Most Astonishing Number* (Hardback ed.). NYC: Broadway (Random House). ISBN 0-7679-0815-5.
- Joseph, George G. (2000) [1991]. *The Crest of the Peacock: The Non-European Roots of Mathematics* (New ed.). Princeton, NJ: Princeton University Press. ISBN 0-691-00659-8.
- Sahlqvist, Leif (2008). *Cardinal Alignments and the Golden Section: Principles of Ancient Cosmography and Design* (3rd Rev. ed.). Charleston, SC: BookSurge. ISBN 1-4196-2157-2.
- Schneider, Michael S. (1994). *A Beginner's Guide to Constructing the Universe: The Mathematical Archetypes of Nature, Art, and Science*. New York: HarperCollins. ISBN 0-06-016939-7.
- Stakhov, A. P. (2009). *The Mathematics of Harmony: From Euclid to Contemporary Mathematics and Computer Science*. Singapore: World Scientific Publishing. ISBN 978-981-277-582-5.
- Walser, Hans (2001) [*Der Goldene Schnitt* 1993]. *The Golden Section*. Peter Hilton trans.. Washington, DC: The Mathematical Association of America. ISBN 0-88385-534-8.
- Scimone, Aldo (1997). *La Sezione Aurea. Storia culturale di un leitmotiv della Matematica*. Palermo: Sigma Edizioni. ISBN 88.7231.025.6.

## External links

- Hazewinkel, Michiel, ed. (2001), "Golden ratio" (<http://www.encyclopediaofmath.org/index.php?title=p/g044570>), *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- "Golden Section" (<http://demonstrations.wolfram.com/GoldenSection/>) by Michael Schreiber, Wolfram Demonstrations Project, 2007.
- Golden Section in Photography: Golden Ratio, Golden Triangles, Golden Spiral ([http://photoinf.com/Golden\\_Mean/Eugene\\_Ilchenko/GoldenSection.html](http://photoinf.com/Golden_Mean/Eugene_Ilchenko/GoldenSection.html))
- Weisstein, Eric W., "Golden Ratio" (<http://mathworld.wolfram.com/GoldenRatio.html>) from MathWorld.
- "Researcher explains mystery of golden ratio" (<http://www.physorg.com/news180531747.html>). *PhysOrg*. December 21, 2009.
- Knott, Ron. "The Golden section ratio: Phi" (<http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/phi.html>). Information and activities by a mathematics professor.
- The Pentagon & The Golden Ratio (<http://web.archive.org/web/20071105084747/http://www.contracosta.ca.us/math/pentagrm.htm>). Green, Thomas M. Updated June 2005. Archived November 2007. Geometry instruction with problems to solve.
- Schneider, Robert P. (2011). "A Golden Pair of Identities in the Theory of Numbers". arXiv:1109.3216 [math.HO]. Proves formulas that involve the golden mean and the Euler totient and Möbius functions.

# Introduction to quantum mechanics

**Quantum mechanics** is an area of physics dealing with phenomena where the action is on the order of the Planck constant. The Planck constant is a very tiny amount and thus this is typically on the distance and momentum scale of atoms. Action is a general physical concept related to dynamics and is most easily recognized in the form of angular momentum. That most tangible way of expressing the essence of quantum mechanics is that we live in a universe of quantized angular momentum and the Planck constant is the quantum. A tangible result of the quantization of angular momentum is the existence of discrete electron orbitals, each with a principal quantum number and each orbital with an associated angular momentum that is a integer multiple of the Planck constant. Quantum mechanics has many implications on the microscopic scale, some of which are obscure and even counterintuitive.

Classical physics explains matter and energy at the macroscopic level of the scale familiar to human experience, including the behavior of astronomical bodies. It remains the key to measurement for much of modern science and technology but at the end of the 19th Century scientists discovered phenomena in both the large (macro) and the small (micro) worlds that classical physics could not explain. Coming to terms with these limitations led to the development of quantum mechanics, a major revolution in physics. This article describes how physicists discovered the limitations of classical physics and developed the main concepts of the quantum theory that replaced them in the early decades of the 20th century.<sup>[1]</sup> These concepts are described in roughly the order they were first discovered; for a more complete history of the subject, see History of quantum mechanics.<sup>[2]</sup>

Some aspects of quantum mechanics can seem counter-intuitive or even paradoxical, because they describe behavior quite different than that seen at larger length scales, where classical physics is an excellent approximation. In the words of Richard Feynman, quantum mechanics deals with "nature as She is — absurd."<sup>[3]</sup>

Many types of energy, such as photons (discrete units of light), behave in some respects like particles and in other respects like waves. Radiators of photons (such as neon lights) have emission spectra that are discontinuous, in that only certain frequencies of light are present. Quantum mechanics predicts the energies, the colours, and the spectral intensities of all forms of electromagnetic radiation.

Quantum mechanics ordains that the more closely one pins down one measure (such as the position of a particle), the less precise another measurement pertaining to the same particle (such as its momentum) must become. This is called the uncertainty principle, also known as the Heisenberg principle after the person who first proposed it.

Put another way, measuring position first and then measuring momentum does *not* have the same outcome as measuring momentum first and then measuring position; the act of measuring the first property necessarily introduces additional energy into the micro-system being studied, thereby perturbing that system.

Even more disconcerting, pairs of particles can be created as "entangled twins." As is described in more detail in the article on Quantum entanglement, entangled particles seem to exhibit what Einstein called "spooky action at a distance," matches between states that classical physics would insist must be random even when distance and the speed of light ensure that no physical causation could account for these correlations.<sup>[4]</sup>



From above and from left to right: Max Planck, Albert Einstein, Niels Bohr, Louis de Broglie, Max Born, Paul Dirac, Werner Heisenberg, Wolfgang Pauli, Erwin Schrödinger, Richard Feynman.

## The first quantum theory: Max Planck and black body radiation



Hot metalwork from a blacksmith. The yellow-orange glow is the visible part of the thermal radiation emitted due to the high temperature. Everything else in the picture is glowing with thermal radiation as well, but less brightly and at longer wavelengths than the human eye can detect. A far-infrared camera can observe this radiation.

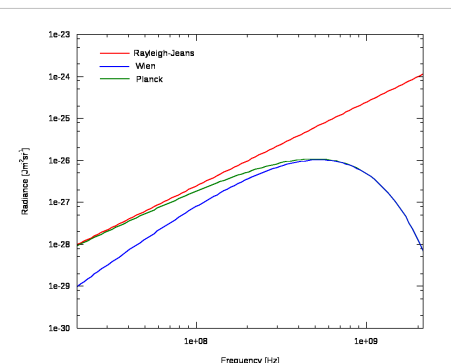
Thermal radiation is electromagnetic radiation emitted from the surface of an object due to the object's temperature. If an object is heated sufficiently, it starts to emit light at the red end of the spectrum — it is *red hot*. Heating it further causes the colour to change from red to yellow to blue to white, as light at shorter wavelengths (higher frequencies) begins to be emitted. It turns out that a perfect emitter is also a perfect absorber. When it is cold, such an object looks perfectly black, because it absorbs all the light that falls on it and emits none. Consequently, an ideal thermal emitter is known as a black body, and the radiation it emits is called black body radiation.

In the late 19th century, thermal radiation had been fairly well-characterized experimentally. How the wavelength at which the radiation is strongest changes with temperature is given by Wien's displacement law, and the overall power emitted per unit area is given by the Stefan–Boltzmann law. However, classical physics was unable to *explain* the relationship between temperatures and predominant frequencies of radiation. In fact, at short wavelengths, classical physics predicted that energy will be emitted by a hot body at an infinite rate. This result, which is clearly wrong, is known as the ultraviolet catastrophe. Physicists were searching for a single theory that explained why they got the experimental results that they did.

The first model that was able to explain the full spectrum of thermal radiation was put forward by Max Planck in 1900.<sup>[5]</sup> He modeled the thermal radiation as being in equilibrium, using a set of harmonic oscillators. To reproduce the experimental results he had to assume that each oscillator produced an integer number of units of energy at its single characteristic frequency, rather than being able to emit any arbitrary amount of energy. In other words, the energy of each oscillator was "quantized."<sup>[6]</sup> The quantum of energy for each oscillator, according to Planck, was proportional to the frequency of the oscillator; the constant of proportionality is now known as the Planck constant. The Planck constant, usually written as  $h$ , has the value  $6.63 \times 10^{-34}$  J s, and so the energy  $E$  of an oscillator of frequency  $f$  is given by

$$E = nhf, \quad \text{where } n = 1, 2, 3, \dots^{[7]}$$

Planck's law was the first quantum theory in physics, and Planck won the Nobel Prize in 1918 "in recognition of the services he rendered to the advancement of Physics by his discovery of energy quanta."<sup>[8]</sup> At the time, however, Planck's view was that quantization was purely a mathematical trick, rather than (as we now know) a fundamental change in our understanding of the world.<sup>[9]</sup>



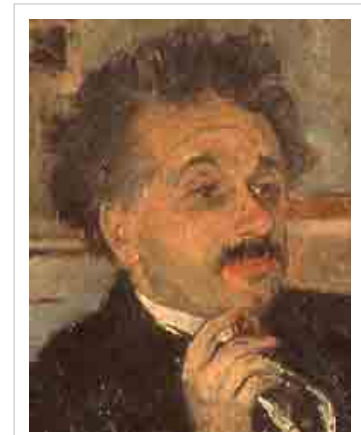
Correct values (green) contrasted against the classical values (Rayleigh-Jeans law, red and Wien approximation, blue).

## Photons: the quantisation of light

In 1905, Albert Einstein took an extra step. He suggested that quantisation was not just a mathematical trick: the energy in a beam of light occurs in individual packets, which are now called photons.<sup>[10]</sup> The energy of a single photon is given by its frequency multiplied by Planck's constant:

$$E = hf.$$

For centuries, scientists had debated between two possible theories of light: was it a wave or did it instead comprise a stream of tiny particles? By the 19th century, the debate was generally considered to have been settled in favour of the wave theory, as it was able to explain observed effects such as refraction, diffraction and polarization. James Clerk Maxwell had shown that electricity, magnetism and light are all manifestations of the same phenomenon: the electromagnetic field. Maxwell's equations, which are the complete set of laws of classical electromagnetism, describe light as waves: a combination of oscillating electric and magnetic fields. Because of the preponderance of evidence in favour of the wave theory, Einstein's ideas were met initially with great skepticism. Eventually, however, the photon model became favoured; one of the most significant pieces of evidence in its favour was its ability to explain several puzzling properties of the photoelectric effect, described in the following section. Nonetheless, the wave analogy remained indispensable for helping to understand other characteristics of light, such as diffraction.



Einstein's portrait by Harm Kamerlingh Onnes at the University of Leiden in 1920

## The photoelectric effect

In 1887 Heinrich Hertz observed that light can eject electrons from metal.<sup>[11]</sup> In 1902 Philipp Lenard discovered that the maximum possible energy of an ejected electron is related to the frequency of the light, not to its *intensity*; if the frequency is too low, no electrons are ejected regardless of the intensity. The lowest frequency of light that causes electrons to be emitted, called the threshold frequency, is different for every metal. This observation is at odds with classical electromagnetism, which predicts that the electron's energy should be proportional to the *intensity* of the radiation.<sup>[12]:24</sup>

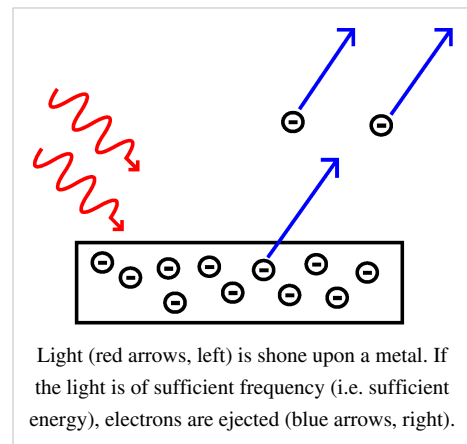
Einstein explained the effect by postulating that a beam of light is a stream of particles (*photons*), and that if the beam is of frequency  $f$  then each photon has an energy equal to  $hf$ .<sup>[11]</sup> An electron is likely to be struck only by a single photon, which imparts at most an energy  $hf$  to the electron.<sup>[11]</sup> Therefore, the intensity of the beam has no effect;<sup>[13]</sup> only its frequency determines the maximum energy that can be imparted to the electron.<sup>[11]</sup>

To explain the threshold effect, Einstein argued that it takes a certain amount of energy, called the *work function*, denoted by  $\varphi$ , to remove an electron from the metal.<sup>[11]</sup> This amount of energy is different for each metal. If the energy of the photon is less than the work function then it does not carry sufficient energy to remove the electron from the metal. The threshold frequency,  $f_0$ , is the frequency of a photon whose energy is equal to the work function:

$$\varphi = hf_0.$$

If  $f$  is greater than  $f_0$ , the energy  $hf$  is enough to remove an electron. The ejected electron has a kinetic energy  $E_K$  which is, at most, equal to the photon's energy minus the energy needed to dislodge the electron from the metal:

$$E_K = hf - \varphi = h(f - f_0).$$



Light (red arrows, left) is shone upon a metal. If the light is of sufficient frequency (i.e. sufficient energy), electrons are ejected (blue arrows, right).



Einstein's description of light as being composed of particles *extended* Planck's notion of quantised energy: a single photon of a given frequency  $f$  delivers an invariant amount of energy  $hf$ . In other words, individual photons can deliver more or less energy, but only depending on their frequencies. However, although the photon is a *particle* it was still being described as having the wave-like property of frequency. Once again, the particle account of light was being "compromised".<sup>[14][15]</sup>

The relationship between the frequency of electromagnetic radiation and the energy of each individual photon is why ultraviolet light can cause sunburn, but visible or infrared light cannot. A photon of ultraviolet light will deliver a high amount of energy—enough to contribute to cellular damage such as occurs in a sunburn. A photon of infrared light will deliver a lower amount of energy—only enough to warm one's skin. So an infrared lamp can warm a large surface, perhaps large enough to keep people comfortable in a cold room, but it cannot give anyone a sunburn.

If each individual photon had identical energy, it would not be correct to talk of a "high energy" photon. Light of high frequency could carry more energy only because of flooding a surface with more photons arriving *per second*. Light of low frequency could carry more energy only for the same reason. If it were true that all photons carry the same energy, then if you doubled the rate of photon delivery, you would double the number of energy units arriving each second. Einstein rejected that wave-dependent classical approach in favour of a particle-based analysis where the energy of the particle must be absolute and varies with frequency in discrete steps (i.e. is quantised). All photons of the same frequency have identical energy, and all photons of different frequencies have proportionally different energies.

In nature, single photons are rarely encountered. The sun emits photons continuously at all electromagnetic frequencies, so they appear to propagate as a continuous wave, not as discrete units. The emission sources available to Hertz and Lennard in the 19th century shared that characteristic. A sun that radiates red light, or a piece of iron in a forge that glows red, may both be said to contain a great deal of energy. It might be surmised that adding continuously to the total energy of some radiating body would make it radiate red light, orange light, yellow light, green light, blue light, violet light, and so on in that order. But that is not so, as larger suns and larger pieces of iron in a forge would glow with colours more toward the violet end of the spectrum. To change the color of such a radiating body it is necessary to change its temperature. An increase in temperature changes the quanta of energy available to excite individual atoms to higher levels, enabling them to emit photons of higher frequencies.

The total energy emitted per unit of time by a sun (or by a piece of iron in a forge) depends on both the number of photons emitted per unit of time, as well as the amount of energy carried by each of the photons involved. In other words, the characteristic frequency of a radiating body is dependent on its temperature. When physicists were looking only at beams of light containing huge numbers of individual and virtually indistinguishable photons, it was difficult to understand the importance of the energy levels of individual photons. So when physicists first discovered devices exhibiting the photoelectric effect, they initially expected that a higher intensity of light would produce a higher voltage from the photoelectric device. Conversely, they discovered that strong beams of light toward the red end of the spectrum might produce no electrical potential at all, and that weak beams of light toward the violet end of the spectrum would produce higher and higher voltages. Einstein's idea that individual units of light may contain different amounts of energy, depending on their frequency, made it possible to explain such experimental results that had hitherto seemed quite counter-intuitive.

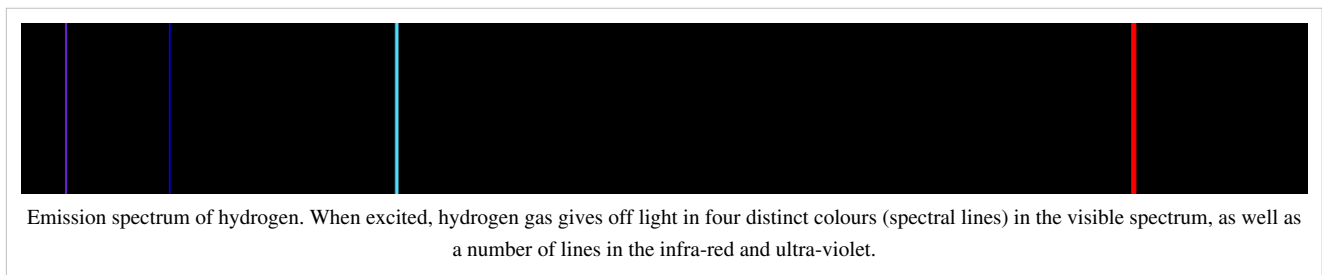
Although the energy imparted by photons is invariant at any given frequency, the initial energy state of the electrons in a photoelectric device prior to absorption of light is not necessarily uniform. Anomalous results may occur in the case of individual electrons. For instance, an electron that was already excited above the equilibrium level of the photoelectric device might be ejected when it absorbed uncharacteristically low frequency illumination. Statistically, however, the characteristic behavior of a photoelectric device will reflect the behavior of the vast majority of its electrons, which will be at their equilibrium level. This point is helpful in comprehending the distinction between the study of individual particles in quantum dynamics and the study of massed particles in classical physics.



## The quantisation of matter: the Bohr model of the atom

By the dawn of the 20th century, it was known that atoms comprise a diffuse cloud of negatively-charged electrons surrounding a small, dense, positively-charged nucleus. This understanding suggested a model in which the electrons circle around the nucleus like planets orbiting a sun.<sup>[16]</sup> However, it was also known that the atom in this model would be unstable: according to classical theory orbiting electrons are undergoing centripetal acceleration, and should therefore give off electromagnetic radiation, the loss of energy also causing them to spiral toward the nucleus, colliding with it in a fraction of a second.

A second, related, puzzle was the emission spectrum of atoms. When a gas is heated, it gives off light only at discrete frequencies. For example, the visible light given off by hydrogen consists of four different colours, as shown in the picture below. By contrast, white light consists of a continuous emission across the whole range of visible frequencies.



Emission spectrum of hydrogen. When excited, hydrogen gas gives off light in four distinct colours (spectral lines) in the visible spectrum, as well as a number of lines in the infra-red and ultra-violet.

In 1885 the Swiss mathematician Johann Balmer discovered that each wavelength  $\lambda$  (lambda) in the visible spectrum of hydrogen is related to some integer  $n$  by the equation

$$\lambda = B \left( \frac{n^2}{n^2 - 4} \right) \quad n = 3, 4, 5, 6$$

where  $B$  is a constant which Balmer determined to be equal to 364.56 nm. Thus Balmer's constant was the basis of a system of discrete, i.e. quantised, integers.

In 1888 Johannes Rydberg generalized and greatly increased the explanatory utility of Balmer's formula. He predicted that  $\lambda$  is related to two integers  $n$  and  $m$  according to what is now known as the Rydberg formula.<sup>[17]</sup>

$$\frac{1}{\lambda} = R \left( \frac{1}{m^2} - \frac{1}{n^2} \right),$$

where  $R$  is the Rydberg constant, equal to  $0.0110 \text{ nm}^{-1}$ , and  $n$  must be greater than  $m$ .

Rydberg's formula accounts for the four visible wavelengths of hydrogen by setting  $m = 2$  and  $n = 3, 4, 5, 6$ . It also predicts additional wavelengths in the emission spectrum: for  $m = 1$  and for  $n > 1$ , the emission spectrum should contain certain ultraviolet wavelengths, and for  $m = 3$  and  $n > 3$ , it should also contain certain infrared wavelengths. Experimental observation of these wavelengths came two decades later: in 1908 Louis Paschen found some of the predicted infrared wavelengths, and in 1914 Theodore Lyman found some of the predicted ultraviolet wavelengths.<sup>[17]</sup>

## Bohr's model

In 1913 Niels Bohr proposed a new model of the atom that included quantized electron orbits.<sup>[18]</sup> In Bohr's model, electrons could inhabit only certain orbits around the atomic nucleus. When an atom emitted (or absorbed) energy, the electron did not move in a continuous trajectory from one orbit around the nucleus to another, as might be expected classically. Instead, the electron would jump instantaneously from one orbit to another, giving off the emitted light in the form of a photon.<sup>[19]</sup> The possible energies of photons given off by each element were determined by the differences in energy between the orbits, and so the emission spectrum for each element would contain a number of lines.<sup>[20]</sup>

Bohr theorised that the angular momentum,  $L$ , of an electron is quantised:

$$L = n \frac{h}{2\pi},$$

where  $n$  is an integer and  $h$  is the Planck constant. Starting from this assumption, Coulomb's law and the equations of circular motion show that an electron with  $n$  units of angular momentum will orbit a proton at a distance  $r$  given by

$$r = \frac{n^2 h^2}{4\pi^2 k_e m e^2},$$

where  $k_e$  is the Coulomb constant,  $m$  is the mass of an electron, and  $e$  is the charge on an electron. For simplicity this is written as

$$r = n^2 a_0,$$

where  $a_0$ , called the Bohr radius, is equal to 0.0529 nm. The Bohr radius is the radius of the smallest allowed orbit.

The energy of the electron<sup>[21]</sup> can also be calculated, and is given by

$$E = -\frac{k_e e^2}{2a_0} \frac{1}{n^2}.$$

Thus Bohr's assumption that angular momentum is quantised means that an electron can only inhabit certain orbits around the nucleus, and that it can have only certain energies. A consequence of these constraints is that the electron will not crash into the nucleus: it cannot continuously emit energy, and it cannot come closer to the nucleus than  $a_0$  (the Bohr radius).

An electron loses energy by jumping instantaneously from its original orbit to a lower orbit; the extra energy is emitted in the form of a photon. Conversely, an electron that absorbs a photon gains energy, hence it jumps to an orbit that is farther from the nucleus.

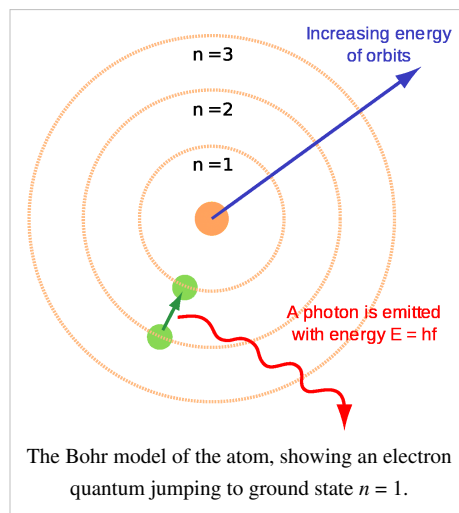
Each photon from glowing atomic hydrogen is due to an electron moving from a higher orbit, with radius  $r_n$ , to a lower orbit,  $r_m$ . The energy  $E_\gamma$  of this photon is the difference in the energies  $E_n$  and  $E_m$  of the electron:

$$E_\gamma = E_n - E_m = \frac{k_e e^2}{2a_0} \left( \frac{1}{m^2} - \frac{1}{n^2} \right)$$

Since Planck's equation shows that the photon's energy is related to its wavelength by  $E_\gamma = hc/\lambda$ , the wavelengths of light that can be emitted are given by

$$\frac{1}{\lambda} = \frac{k_e e^2}{2a_0 hc} \left( \frac{1}{m^2} - \frac{1}{n^2} \right).$$

This equation has the same form as the Rydberg formula, and predicts that the constant  $R$  should be given by



$$R = \frac{k_e e^2}{2a_0 h c}.$$

Therefore the Bohr model of the atom can predict the emission spectrum of hydrogen in terms of fundamental constants.<sup>[22]</sup> However, it was not able to make accurate predictions for multi-electron atoms, or to explain why some spectral lines are brighter than others.

## Wave-particle duality

In 1924, Louis de Broglie proposed the idea that just as light has both wave-like and particle-like properties, matter also has wave-like properties.<sup>[23]</sup>

The wavelength,  $\lambda$ , associated with a particle is related to its momentum,  $p$  through the Planck constant  $h$ :<sup>[24][25]</sup>

$$p = \frac{h}{\lambda}.$$

The relationship, called the de Broglie hypothesis, holds for all types of matter. Thus all matter exhibits properties of both particles and waves.

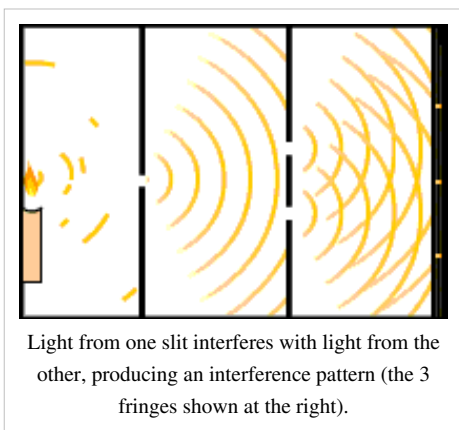
Three years later, the wave-like nature of electrons was demonstrated by showing that a beam of electrons could exhibit diffraction, just like a beam of light. At the University of Aberdeen, George Thomson passed a beam of electrons through a thin metal film and observed the predicted diffraction patterns. At Bell Labs, Davisson and Germer guided their beam through a crystalline grid. Similar wave-like phenomena were later shown for atoms and even small molecules. De Broglie was awarded the Nobel Prize for Physics in 1929 for his hypothesis; Thomson and Davisson shared the Nobel Prize for Physics in 1937 for their experimental work.

The concept of wave-particle duality says that neither the classical concept of "particle" nor of "wave" can fully describe the behavior of quantum-scale objects, either photons or matter. Indeed, astrophysicist A.S. Eddington proposed in 1927 that "We can scarcely describe such an entity as a wave or as a particle; perhaps as a compromise we had better call it a 'wavicle' ".<sup>[26]</sup> (This term was later popularised by mathematician Banesh Hoffmann.)<sup>[27]:172</sup>

Wave-particle duality is an example of the principle of complementarity in quantum physics. An elegant example of wave-particle duality, the double slit experiment, is discussed in the section below.

De Broglie's treatment of quantum events served as a jumping off point for Schrödinger when he set about to construct a wave equation to describe quantum theoretical events.

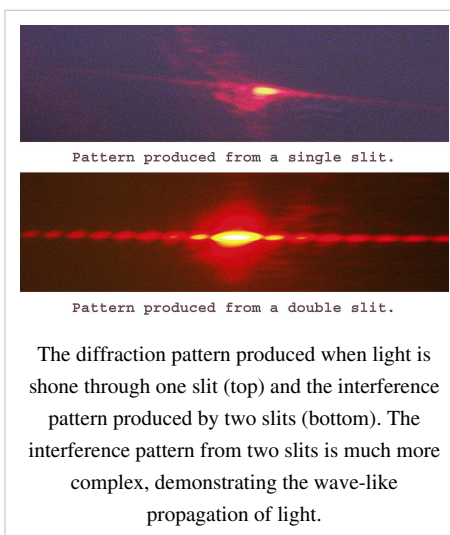
## The double-slit experiment



In the double-slit experiment as originally performed by Thomas Young and Augustin Fresnel in 1827, a beam of light is directed through two narrow, closely spaced slits, producing an interference pattern of light and dark bands on a screen. If one of the slits is covered up, one might naively expect that the intensity of the fringes due to interference would be halved everywhere. In fact, a much simpler pattern is seen, a simple diffraction pattern. Closing one slit results in a much simpler pattern diametrically opposite the open slit. Exactly the same behaviour can be demonstrated in water waves, and so the double-slit experiment was seen as a demonstration of the wave nature of light.

The double-slit experiment has also been performed using electrons, atoms, and even molecules, and the same type of interference pattern is seen. Thus it has been demonstrated that all matter possesses both particle and wave characteristics.

Even if the source intensity is turned down so that only one particle (e.g. photon or electron) is passing through the apparatus at a time, the same interference pattern develops over time. The quantum particle acts as a wave when passing through the double slits, but as a particle when it is detected. This is a typical feature of quantum complementarity: a quantum particle will act as a wave when we do an experiment to measure its wave-like properties, and like a particle when we do an experiment to measure its particle-like properties. Where on the detector screen any individual particle shows up will be the result of an entirely random process.



### Application to the Bohr model

De Broglie expanded the Bohr model of the atom by showing that an electron in orbit around a nucleus could be thought of as having wave-like properties. In particular, an electron will be observed only in situations that permit a standing wave around a nucleus. An example of a standing wave is a violin string, which is fixed at both ends and can be made to vibrate. The waves created by a stringed instrument appear to oscillate in place, moving from crest to trough in an up-and-down motion. The wavelength of a standing wave is related to the length of the vibrating object and the boundary conditions. For example, because the violin string is fixed at both ends, it can carry standing waves of wavelengths  $2l/n$ , where  $l$  is the length and  $n$  is a positive integer. De Broglie suggested that the allowed electron orbits were those for which the circumference of the orbit would be an integer number of wavelengths.

### Development of modern quantum mechanics

In 1925, building on de Broglie's hypothesis, Erwin Schrödinger developed the equation that describes the behaviour of a quantum mechanical wave. The equation, called the Schrödinger equation after its creator, is central to quantum mechanics, defines the permitted stationary states of a quantum system, and describes how the quantum state of a physical system changes in time.<sup>[28]</sup> In the paper that introduced Schrödinger's cat, he says that the psi-function featured in his equation provides the "means for predicting probability of measurement results," and that it therefore provides "future expectation[s] , somewhat as laid down in a *catalog*."<sup>[29]</sup>

Schrödinger was able to calculate the energy levels of hydrogen by treating a hydrogen atom's electron as a classical wave, moving in a well of electrical potential created by the proton. This calculation accurately reproduced the energy levels of the Bohr model.

At a somewhat earlier time, Werner Heisenberg was trying to find an explanation for the intensities of the different lines in the hydrogen emission spectrum. By means of a series of mathematical analogies, Heisenberg wrote out the quantum mechanical analogue for the classical computation of intensities. Shortly afterwards, Heisenberg's colleague Max Born realised that Heisenberg's method of calculating the



probabilities for transitions between the different energy levels could best be expressed by using the mathematical concept of matrices.<sup>[30]</sup>

In May 1926, Schrödinger proved that Heisenberg's matrix mechanics and his own wave mechanics made the same predictions about the properties and behaviour of the electron; mathematically, the two theories were identical. Yet the two men disagreed on the interpretation of their mutual theory. For instance, Heisenberg saw no problem in the theoretical prediction of instantaneous transitions of electrons between orbits in an atom, but Schrödinger hoped that a theory based on continuous wave-like properties could avoid what he called (in the words of Wilhelm Wien<sup>[31]</sup>) "this nonsense about quantum jumps."

## Copenhagen interpretation

Bohr, Heisenberg and others tried to explain what these experimental results and mathematical models really mean. Their description, known as the Copenhagen interpretation of quantum mechanics, aimed to describe the nature of reality that was being probed by the measurements and described by the mathematical formulations of quantum mechanics.

The main principles of the Copenhagen interpretation are:

1. A system is completely described by a wave function,  $\psi$ . (Heisenberg)
2. How  $\psi$  changes over time is given by the Schrödinger equation.
3. The description of nature is essentially probabilistic. The probability of an event — for example, where on the screen a particle will show up in the two slit experiment — is related to the square of the amplitude of its wave function. (Born rule, due to Max Born, which gives a physical meaning to the wavefunction in the Copenhagen interpretation: the probability amplitude)
4. It is not possible to know the values of all of the properties of the system at the same time; those properties that are not known with precision must be described by probabilities. (Heisenberg's uncertainty principle)
5. Matter, like energy, exhibits a wave-particle duality. An experiment can demonstrate the particle-like properties of matter, or its wave-like properties; but not both at the same time. (Complementarity principle due to Bohr)
6. Measuring devices are essentially classical devices, and measure classical properties such as position and momentum.
7. The quantum mechanical description of large systems should closely approximate the classical description. (Correspondence principle of Bohr and Heisenberg)

Various consequences of these principles are discussed in more detail in the following subsections.

## Uncertainty principle

Suppose that we want to measure the position and speed of an object — for example a car going through a radar speed trap. Naively, we assume that the car has a definite position and speed at a particular moment in time, and how accurately we can measure these values depends on the quality of our measuring equipment — if we improve the precision of our measuring equipment, we will get a result that is closer to the true value. In particular, we would assume that how precisely we measure the speed of the car does not affect the measurement of its position, and vice versa.

In 1927, Heisenberg proved that these assumptions are not correct.<sup>[33]</sup> Quantum mechanics shows that certain pairs of physical properties, like position and speed, cannot both be known to arbitrary precision: the more precisely one property is known, the less precisely the other can be known. This statement is known as the uncertainty principle. The uncertainty principle isn't a statement about the accuracy of our measuring equipment, but about the nature of the system itself — our naive assumption that the car had a definite position and speed was incorrect. On a scale of cars and people, these uncertainties are too small to notice, but when dealing with atoms and electrons they become critical.<sup>[34]</sup>

Heisenberg gave, as an illustration, the measurement of the position and momentum of an electron using a photon of light. In measuring the electron's position, the higher the frequency of the photon the more accurate is the measurement of the position of the impact, but the greater is the disturbance of the electron, which absorbs a random amount of energy, rendering the measurement obtained of its momentum increasingly uncertain (momentum is velocity multiplied by mass), for one is necessarily measuring its post-impact disturbed momentum, from the collision products, not its original momentum. With a photon of lower frequency the disturbance - hence uncertainty - in the momentum is less, but so is the accuracy of the measurement of the position of the impact.<sup>[35]</sup>

The uncertainty principle shows mathematically that the product of the uncertainty in the position and momentum of a particle (momentum is velocity multiplied by mass) could never be less than a certain value, and that this value is related to Planck's constant.

## Wave function collapse

Wave function collapse is a forced term for whatever happened when it becomes appropriate to replace the description of an uncertain state of a system by a description of the system in a definite state. Explanations for the nature of the process of becoming certain are controversial. At any time before a photon "shows up" on a detection screen it can only be described by a set of probabilities for where it might show up. When it does show up, for instance in the CCD of an electronic camera, the time and the space where it interacted with the device are known within very tight limits. However, the photon has disappeared, and the wave function has disappeared with it. In its place some physical change in the detection screen has appeared, e.g., an exposed spot in a sheet of photographic film, or a change in electric potential in some cell of a CCD.



Werner Heisenberg at the age of 26. Heisenberg won the Nobel Prize in Physics in 1932 for the work that he did at around this time.<sup>[32]</sup>

## Eigenstates and eigenvalues

*For a more detailed introduction to this subject, see: Introduction to eigenstates*

Because of the uncertainty principle, statements about both the position and momentum of particles can only assign a probability that the position or momentum will have some numerical value. Therefore it is necessary to formulate clearly the difference between the state of something that is indeterminate, such as an electron in a probability cloud, and the state of something having a definite value. When an object can definitely be "pinned-down" in some respect, it is said to possess an eigenstate.

## The Pauli exclusion principle

In 1924, Wolfgang Pauli proposed a new quantum degree of freedom (or quantum number), with two possible values, to resolve inconsistencies between observed molecular spectra and the predictions of quantum mechanics. In particular, the spectrum of atomic hydrogen had a doublet, or pair of lines differing by a small amount, where only one line was expected. Pauli formulated his *exclusion principle*, stating that "There cannot exist an atom in such a quantum state that two electrons within [it] have the same set of quantum numbers."<sup>[36]</sup>

A year later, Uhlenbeck and Goudsmit identified Pauli's new degree of freedom with a property called spin. The idea, originating with Ralph Kronig, was that electrons behave as if they rotate, or "spin", about an axis. Spin would account for the missing magnetic moment, and allow two electrons in the same orbital to occupy distinct quantum states if they "spun" in opposite directions, thus satisfying the exclusion principle. The quantum number represented the sense (positive or negative) of spin.

## Application to the hydrogen atom

Bohr's model of the atom was essentially two-dimensional — an electron orbiting in a plane around its nuclear "sun." However, the uncertainty principle states that an electron cannot be viewed as having an exact location at any given time. In the modern theory the orbit has been replaced by an *atomic orbital*, a "cloud" of possible locations. It is often depicted as a three-dimensional region within which there is a 95 percent probability of finding the electron.<sup>[37]</sup>

Schrödinger was able to calculate the energy levels of hydrogen by treating a hydrogen atom's electron as a wave, represented by the "wave function"  $\Psi$ , in a electric potential well,  $V$ , created by the proton. The solutions to Schrödinger's equation are distributions of probabilities for electron positions and locations. Orbitals have a range of different shapes in three dimensions. The energies of the different orbitals can be calculated, and they accurately reproduce the energy levels of the Bohr model.

Within Schrödinger's picture, each electron has four properties:

1. An "orbital" designation, indicating whether the particle wave is one that is closer to the nucleus with less energy or one that is farther from the nucleus with more energy;
2. The "shape" of the orbital, spherical or otherwise;
3. The "inclination" of the orbital, determining the magnetic moment of the orbital around the  $z$ -axis.
4. The "spin" of the electron.

The collective name for these properties is the quantum state of the electron. The quantum state can be described by giving a number to each of these properties; these are known as the electron's quantum numbers. The quantum state of the electron is described by its wavefunction. The Pauli exclusion principle demands that no two electrons within an atom may have the same values of all four numbers.



The first property describing the orbital is the principal quantum number,  $n$ , which is the same as in Bohr's model.  $n$  denotes the energy level of each orbital. The possible values for  $n$  are integers:

$$n = 1, 2, 3 \dots$$

The next quantum number, the azimuthal quantum number, denoted  $l$ , describes the shape of the orbital. The shape is a consequence of the angular momentum of the orbital. The angular momentum represents the resistance of a spinning object to speeding up or slowing down under the influence of external force. The azimuthal quantum number represents the orbital angular momentum of an electron around its nucleus. The possible values for  $l$  are integers from 0 to  $n - 1$ :

$$l = 0, 1, \dots, n - 1.$$

The shape of each orbital has its own letter as well. The first shape is denoted by the letter  $s$  (a mnemonic being "sphere"). The next shape is denoted by the letter  $p$  and has the form of a dumbbell. The other orbitals have more complicated shapes (see atomic orbital), and are denoted by the letters  $d$ ,  $f$ , and  $g$ .

The third quantum number, the magnetic quantum number, describes the magnetic moment of the electron, and is denoted by  $m_l$  (or simply  $m$ ). The possible values for  $m_l$  are integers from  $-l$  to  $l$ :

$$m_l = -l, -(l - 1), \dots, 0, 1, \dots, l.$$

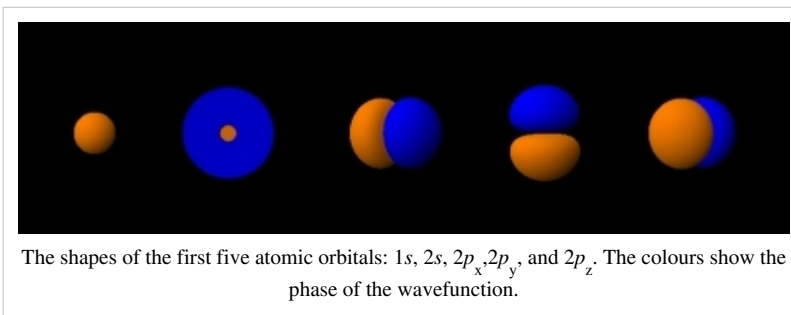
The magnetic quantum number measures the component of the angular momentum in a particular direction. The choice of direction is arbitrary, conventionally the  $z$ -direction is chosen.

The fourth quantum number, the spin quantum number (pertaining to the "orientation" of the electron's spin) is denoted  $m_s$ , with values  $+\frac{1}{2}$  or  $-\frac{1}{2}$ .

The chemist Linus Pauling wrote, by way of example:

In the case of a helium atom with two electrons in the  $1s$  orbital, the Pauli Exclusion Principle requires that the two electrons differ in the value of one quantum number. Their values of  $n$ ,  $l$ , and  $m_l$  are the same; moreover, they have the same spin,  $s = \frac{1}{2}$ . Accordingly they must differ in the value of  $m_s$ , which can have the value of  $+\frac{1}{2}$  for one electron and  $-\frac{1}{2}$  for the other."<sup>[36]</sup>

It is the underlying structure and symmetry of atomic orbitals, and the way that electrons fill them, that determines the organisation of the periodic table and the structure and strength of chemical bonds between atoms.

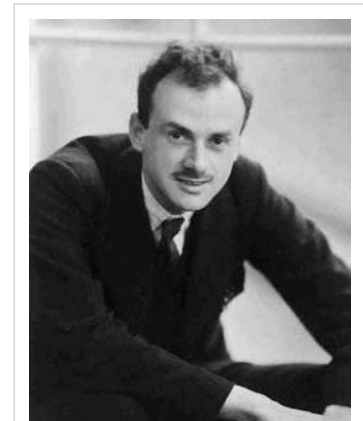


The shapes of the first five atomic orbitals:  $1s$ ,  $2s$ ,  $2p_x$ ,  $2p_y$ , and  $2p_z$ . The colours show the phase of the wavefunction.

## Dirac wave equation

In 1928, Paul Dirac extended the Pauli equation, which described spinning electrons, to account for special relativity. The result was a theory that dealt properly with events, such as the speed at which an electron orbits the nucleus, occurring at a substantial fraction of the speed of light. By using the simplest electromagnetic interaction, Dirac was able to predict the value of the magnetic moment associated with the electron's spin, and found the experimentally observed value, which was too large to be that of a spinning charged sphere governed by classical physics. He was able to solve for the spectral lines of the hydrogen atom, and to reproduce from physical first principles Sommerfeld's successful formula for the fine structure of the hydrogen spectrum.

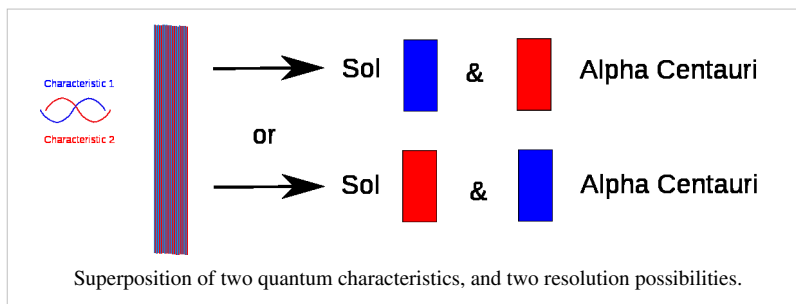
Dirac's equations sometimes yielded a negative value for energy, for which he proposed a novel solution: he posited the existence of an antielectron and of a dynamical vacuum. This led to the many-particle quantum field theory.



Paul Dirac (1902 - 1984)

$$(c\boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + \beta mc^2) \psi = i\hbar \frac{\partial \psi}{\partial t}$$

## Quantum entanglement



The Pauli exclusion principle says that two electrons in one system cannot be in the same state. Nature leaves open the possibility, however, that two electrons can have both states "superimposed" over each of them. Recall that the wave functions that emerge simultaneously from the double slits arrive at the detection

screen in a state of superposition. Nothing is certain until the superimposed waveforms "collapse," At that instant an electron shows up somewhere in accordance with the probabilities that are the squares of the amplitudes of the two superimposed waveforms. The situation there is already very abstract. A concrete way of thinking about entangled photons, photons in which two contrary states are superimposed on each of them in the same event, is as follows:

Imagine that the superposition of a state that can be mentally **labeled** as blue and another state that can be mentally labeled as red will then appear (in imagination, of course) as a purple state. Two photons are produced as the result of the same atomic event. Perhaps they are produced by the excitation of a crystal that characteristically absorbs a photon of a certain frequency and emits two photons of half the original frequency. So the two photons come out "purple." If the experimenter now performs some experiment that will determine whether one of the photons is either blue or red, then that experiment changes the photon involved from one having a superposition of "blue" and "red" characteristics to a photon that has only one of those characteristics. The problem that Einstein had with such an imagined situation was that if one of these photons had been kept bouncing between mirrors in a laboratory on earth, and the other one had traveled halfway to the nearest star, when its twin was made to reveal itself as either blue or red, that meant that the distant photon now had to lose its "purple" status too. So whenever it might be investigated after its twin had been measured, it would necessarily show up in the opposite state to whatever its twin had revealed.

In trying to show that quantum mechanics was not a complete theory, Einstein started with the theory's prediction that two or more particles that have interacted in the past can appear strongly correlated when their various properties are later measured. He sought to explain this seeming interaction in a classical way, through their common past, and preferably not by some "spooky action at a distance." The argument is worked out in a famous paper, Einstein, Podolsky, and Rosen (1935; abbreviated EPR), setting out what is now called the EPR paradox. Assuming what is now usually called local realism, EPR attempted to show from quantum theory that a particle has both position and momentum simultaneously, while according to the Copenhagen interpretation, only one of those two properties actually exists and only at the moment that it is being measured. EPR concluded that quantum theory is incomplete in that it refuses to consider physical properties which objectively exist in nature. (Einstein, Podolsky, & Rosen 1935 is currently Einstein's most cited publication in physics journals.) In the same year, Erwin Schrödinger used the word "entanglement" and declared: "I would not call that *one* but rather *the* characteristic trait of quantum mechanics."<sup>[38]</sup> The question of whether entanglement is a real condition is still in dispute.<sup>[39]</sup> The Bell inequalities are the most powerful challenge to Einstein's claims.

## Quantum field theory

The idea of quantum field theory began in the late 1920s with British physicist Paul Dirac, when he attempted to quantise the electromagnetic field — a procedure for constructing a quantum theory starting from a classical theory.

A *field* in physics is "a region or space in which a given effect (such as magnetism) exists."<sup>[40]</sup> Other effects that manifest themselves as fields are gravitation and static electricity.<sup>[41]</sup> In 2008, physicist Richard Hammond wrote that

Sometimes we distinguish between quantum mechanics (QM) and quantum field theory (QFT). QM refers to a system in which the number of particles is fixed, and the fields (such as the electromechanical field) are continuous classical entities. QFT . . . goes a step further and allows for the creation and annihilation of particles . . .

He added, however, that *quantum mechanics* is often used to refer to "the entire notion of quantum view."<sup>[42]:108</sup>

In 1931, Dirac proposed the existence of particles that later became known as anti-matter.<sup>[43]</sup> Dirac shared the Nobel Prize in physics for 1933 with Schrödinger, "for the discovery of new productive forms of atomic theory."<sup>[44]</sup>

## Quantum electrodynamics

Quantum electrodynamics (QED) is the name of the quantum theory of the electromagnetic force. Understanding QED begins with understanding electromagnetism. Electromagnetism can be called "electrodynamics" because it is a dynamic interaction between electrical and magnetic forces. Electromagnetism begins with the electric charge.

Electric charges are the sources of, and create, electric fields. An electric field is a field which exerts a force on any particles that carry electric charges, at any point in space. This includes the electron, proton, and even quarks, among others. As a force is exerted, electric charges move, a current flows and a magnetic field is produced. The magnetic field, in turn causes electric current (moving electrons). The interacting electric and magnetic field is called an electromagnetic field.



This sculpture in Bristol, England — a series of clustering cones — presents the idea of small worlds that Paul Dirac studied to reach his discovery of anti-matter.

The physical description of interacting charged particles, electrical currents, electrical fields, and magnetic fields is called electromagnetism.

In 1928 Paul Dirac produced a relativistic quantum theory of electromagnetism. This was the progenitor to modern quantum electrodynamics, in that it had essential ingredients of the modern theory. However, the problem of unsolvable infinities developed in this relativistic quantum theory. Years later, renormalization solved this problem. Initially viewed as a suspect, provisional procedure by some of its originators, renormalization eventually was embraced as an important and self-consistent tool in QED and other fields of physics. Also, in the late 1940s Feynman's diagrams depicted all possible interactions pertaining to a given event. The diagrams showed that the electromagnetic force is the interactions of photons between interacting particles.

An example of a prediction of quantum electrodynamics which has been verified experimentally is the Lamb shift. This refers to an effect whereby the quantum nature of the electromagnetic field causes the energy levels in an atom or ion to deviate slightly from what they would otherwise be. As a result, spectral lines may shift or split.

In the 1960s physicists realized that QED broke down at extremely high energies. From this inconsistency the Standard Model of particle physics was discovered, which remedied the higher energy breakdown in theory. The Standard Model unifies the electromagnetic and weak interactions into one theory. This is called the electroweak theory.

## Interpretations

The physical measurements, equations, and predictions pertinent to quantum mechanics are all consistent and hold a very high level of confirmation. However, the question of what these abstract models say about the underlying nature of the real world has received competing answers.

## Applications

Applications of quantum mechanics include the laser, the transistor, the electron microscope, and magnetic resonance imaging. A special class of quantum mechanical applications is related to macroscopic quantum phenomena such as superfluid helium and superconductors. The study of semiconductors led to the invention of the diode and the transistor, which are indispensable for modern electronics.

In even the simple light switch, quantum tunnelling is absolutely vital, as otherwise the electrons in the electric current could not penetrate the potential barrier made up of a layer of oxide. Flash memory chips found in USB drives also use quantum tunnelling, to erase their memory cells.<sup>[45]</sup>

## Notes

- [1] Classical physics also does not accurately describe the universe on the largest scales or at speeds close to that of light. An accurate description requires general relativity.
- [2] *Quantum Mechanics* from [[National Public Radio (<http://www.pbs.org/transistor/science/info/quantum.html>)]]
- [3] Feynman, Richard P. (1988). *QED : the strange theory of light and matter* (1st Princeton pbk., seventh printing with corrections. ed.). Princeton, N.J.: Princeton University Press. pp. 10. ISBN 978-0691024172.
- [4] Alan Macdonald, "Spooky action at a distance: The puzzle of entanglement in quantum theory," page 5 of 7, downloaded 13 June 2012 from <http://faculty.luther.edu/~macdonal/>
- [5] This result was published (in German) as Planck, Max (1901). "Ueber das Gesetz der Energieverteilung im Normalspectrum" ([http://www.physik.uni-augsburg.de/annalen/history/historic-papers/1901\\_309\\_553-563.pdf](http://www.physik.uni-augsburg.de/annalen/history/historic-papers/1901_309_553-563.pdf)). *Ann. Phys.* **309** (3): 553–63. Bibcode 1901AnP...309..553P. doi:10.1002/andp.19013090310. . English translation: " On the Law of Distribution of Energy in the Normal Spectrum (<http://dbhs.wvusd.k12.ca.us/webdocs/Chem-History/Planck-1901/Planck-1901.html>)".
- [6] The word "quantum" comes from the Latin word for "how much" (as does "quantity"). Something which is "quantized," like the energy of Planck's harmonic oscillators, can only take specific values. For example, in most countries money is effectively quantized, with the "quantum of money" being the lowest-value coin in circulation. "Mechanics" is the branch of science that deals with the action of forces on objects, so "quantum mechanics" is the part of mechanics that deals with objects for which particular properties are quantized.

- [7] Francis Weston Sears (1958). *Mechanics, Wave Motion, and Heat* ([http://books.google.com/books?hl=en&q="Mechanics,+Wave+Motion,+and+Heat"+"where+n+=+1,"&btnG=Search+Books](http://books.google.com/books?hl=en&q=)). Addison-Wesley. p. 537. .
- [8] "The Nobel Prize in Physics 1918" ([http://nobelprize.org/nobel\\_prizes/physics/laureates/1918/](http://nobelprize.org/nobel_prizes/physics/laureates/1918/)). The Nobel Foundation. . Retrieved 2009-08-01.
- [9] Kragh, Helge (1 December 2000). "Max Planck: the reluctant revolutionary" (<http://physicsworld.com/cws/article/print/373>). PhysicsWorld.com.
- [10] Einstein, Albert (1905). "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt" (<http://www.zbp.univie.ac.at/dokumente/einstein1.pdf>). *Annalen der Physik* **17** (6): 132–148. Bibcode 1905AnP...322..132E. doi:10.1002/andp.19053220607. ., translated into English as On a Heuristic Viewpoint Concerning the Production and Transformation of Light ([http://lorenz.phl.jhu.edu/AnnusMirabilis/AeReserveArticles/eins\\_lq.pdf](http://lorenz.phl.jhu.edu/AnnusMirabilis/AeReserveArticles/eins_lq.pdf)). The term "photon" was introduced in 1926.
- [11] Taylor, J. R.; Zafiratos, C. D.; Dubson, M. A. (2004). *Modern Physics for Scientists and Engineers*. Prentice Hall. pp. 127–9. ISBN 0-13-589789-0.
- [12] Stephen Hawking, *The Universe in a Nutshell*, Bantam, 2001.
- [13] Actually there can be intensity-dependent effects, but at intensities achievable with non-laser sources these effects are unobservable.
- [14] Dicke and Wittke, *Introduction to Quantum Mechanics*, p. 12
- [15] Einstein's photoelectric effect equation *can* be derived and explained *without* requiring the concept of "photons". That is, the electromagnetic radiation can be treated as a classical electromagnetic wave, as long as the electrons in the material are treated by the laws of quantum mechanics. The results are quantitatively correct for thermal light sources (the sun, incandescent lamps, etc) both for the rate of electron emission as well as their angular distribution. For more on this point, see NTRS.NASA.gov ([http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680009569\\_1968009569.pdf](http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680009569_1968009569.pdf))
- [16] The classical model of the atom is called the planetary model, or sometimes the Rutherford model after Ernest Rutherford who proposed it in 1911, based on the Geiger-Marsden gold foil experiment which first demonstrated the existence of the nucleus.
- [17] Taylor, J. R.; Zafiratos, C. D.; Dubson, M. A. (2004). *Modern Physics for Scientists and Engineers*. Prentice Hall. pp. 147–8. ISBN 0-13-589789-0.
- [18] McEvoy, J. P.; Zarate, O. (2004). *Introducing Quantum Theory*. Totem Books. pp. 70–89, especially p. 89. ISBN 1-84046-577-8.
- [19] *World Book Encyclopedia*, page 6, 2007.
- [20] Dicke and Wittke, *Introduction to Quantum Mechanics*, p. 10f.
- [21] In this case, the energy of the electron is the sum of its kinetic and potential energies. The electron has kinetic energy by virtue of its actual motion around the nucleus, and potential energy because of its electromagnetic interaction with the nucleus.
- [22] The model can be easily modified to account of the emission spectrum of any system consisting of a nucleus and a single electron (that is, ions such as He<sup>+</sup> or O<sup>7+</sup> which contain only one electron).
- [23] J. P. McEvoy and Oscar Zarate (2004). *Introducing Quantum Theory*. Totem Books. p. 110f. ISBN 1-84046-577-8.
- [24] Aezel, Amir D., *Entanglrmnt*, p. 51f. (Penguin, 2003) ISBN 0-452-28457
- [25] J. P. McEvoy and Oscar Zarate (2004). *Introducing Quantum Theory*. Totem Books. p. 114. ISBN 1-84046-577-8.
- [26] A.S. Eddington, *The Nature of the Physical World*, the course of Gifford Lectures that Eddington delivered in the University of Edinburgh in January to March 1927, Kessinger Publishing, 2005, p. 201. ([http://books.google.com/books?id=PGOTKcxSqMUC&pg=PA201&lpg=PA201&dq=We+can+scarcely+describe+such+an+entity+as+a+wave+or+as+a+particle;+perhaps+as+a+compromise+we+had+better+call+it+a+`wavicl&source=bl&ots=K0lfGzaXli&sig=zgrQiBjBhRLuUzVBT-yy8jZhC1Y&hl=en&ei=i8g1SpOHC4PgtgOu\\_4jVDg&sa=X&oi=book\\_result&ct=result&resnum=1](http://books.google.com/books?id=PGOTKcxSqMUC&pg=PA201&lpg=PA201&dq=We+can+scarcely+describe+such+an+entity+as+a+wave+or+as+a+particle;+perhaps+as+a+compromise+we+had+better+call+it+a+`wavicl&source=bl&ots=K0lfGzaXli&sig=zgrQiBjBhRLuUzVBT-yy8jZhC1Y&hl=en&ei=i8g1SpOHC4PgtgOu_4jVDg&sa=X&oi=book_result&ct=result&resnum=1))
- [27] Banesh Hoffman, *The Strange Story of the Quantum*, Dover, 1959
- [28] "Schrodinger Equation (Physics)," *Encyclopædia Britannica* (<http://www.britannica.com/EBchecked/topic/528298/Schrodinger-equation>)
- [29] Erwin Schrödinger, "The Present Situation in Quantum Mechanics," p. 9. "This translation was originally published in Proceedings of the American Philosophical Society, 124, 323-38. [And then appeared as Section I.11 of Part I of Quantum Theory and Measurement (J.A. Wheeler and W.H. Zurek, eds., Princeton university Press, New Jersey 1983). This paper can be downloaded from <http://www.tu-harburg.de/rzt/rzt/it/QM/cat.html>."
- [30] For a somewhat more sophisticated look at how Heisenberg transitioned from the old quantum theory and classical physics to the new quantum mechanics, see Heisenberg's entryway to matrix mechanics.
- [31] W. Moore, *Schrödinger: Life and Thought*, Cambridge University Press (1989), p. 222.
- [32] Heisenberg's Nobel Prize citation ([http://nobelprize.org/nobel\\_prizes/physics/laureates/1932/](http://nobelprize.org/nobel_prizes/physics/laureates/1932/))
- [33] Heisenberg first published his work on the uncertainty principle in the leading German physics journal *Zeitschrift für Physik*: Heisenberg, W. (1927). "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik". *Z. Phys.* **43** (3–4): 172–198. Bibcode 1927ZPhy...43..172H. doi:10.1007/BF01397280.
- [34] Nobel Prize in Physics presentation speech, 1932 ([http://nobelprize.org/nobel\\_prizes/physics/laureates/1932/press.html](http://nobelprize.org/nobel_prizes/physics/laureates/1932/press.html))
- [35] "Uncertainty principle," *Encyclopædia Britannica* (<http://www.britannica.com/EBchecked/topic/614029/uncertainty-principle>)
- [36] Linus Pauling, **The Nature of the Chemical Bond**, p. 47
- [37] "Orbital (chemistry and physics)," *Encyclopædia Britannica* (<http://www.britannica.com/EBchecked/topic/431159/orbital>)
- [38] E. Schrödinger, *Proceedings of the Cambridge Philosophical Society*, 31 (1935), p. 555says: "When two systems, of which we know the states by their respective representation, enter into a temporary physical interaction due to known forces between them and when after a time

of mutual influence the systems separate again, then they can no longer be described as before, viz., by endowing each of them with a representative of its own. I would not call that *one* but rather *the* characteristic trait of quantum mechanics."

- [39] "Quantum Nonlocality and the Possibility of Superluminal Effects", John G. Cramer, npl.washington.edu ([http://www.npl.washington.edu/npl/int\\_rep/qm\\_nl.html](http://www.npl.washington.edu/npl/int_rep/qm_nl.html))
- [40] "Mechanics," *Merriam-Webster Online Dictionary* (<http://www.merriam-webster.com/dictionary/field>)
- [41] "Field," *Encyclopaedia Britannica* (<http://www.britannica.com/EBchecked/topic/206162/field>)
- [42] Richard Hammond, *The Unknown Universe*, New Page Books, 2008. ISBN 978-1-60163-003-2
- [43] The Physical World website ([http://www.physicalworld.org/restless\\_universe/html/ru\\_dira.html](http://www.physicalworld.org/restless_universe/html/ru_dira.html))
- [44] "The Nobel Prize in Physics 1933" ([http://nobelprize.org/nobel\\_prizes/physics/laureates/1933/](http://nobelprize.org/nobel_prizes/physics/laureates/1933/)). The Nobel Foundation. . Retrieved 2007-11-24.
- [45] Durrani, Z. A. K.; Ahmed, H. (2008). Vijay Kumar. ed. *Nanosilicon*. Elsevier. p. 345. ISBN 978-0-08-044528-1.

## References

- Bernstein, Jeremy (2005). "Max Born and the quantum theory". *American Journal of Physics* **73** (11).
- Beller, Mara (2001). *Quantum Dialogue: The Making of a Revolution*. University of Chicago Press.
- Bohr, Niels (1958). *Atomic Physics and Human Knowledge*. John Wiley & Sons. ASIN B00005VGVF. ISBN 0-486-47928-5. OCLC 530611.
- de Broglie, Louis (1953). *The Revolution in Physics*. Noonday Press. LCCN 53010401.
- Einstein, Albert (1934). *Essays in Science*. Philosophical Library. ISBN 0-486-47011-3. LCCN 55003947.
- Feigl, Herbert; Brodbeck, May (1953). *Readings in the Philosophy of Science*. Appleton-Century-Crofts. ISBN 0-390-30488-3. LCCN 53006438.
- Feynman, Richard P. (1949). "Space-Time Approach to Quantum Electrodynamics" ([http://www.physics.princeton.edu/~mcdonald/examples/QED/feynman\\_pr\\_76\\_769\\_49.pdf](http://www.physics.princeton.edu/~mcdonald/examples/QED/feynman_pr_76_769_49.pdf)). *Physical Review* **76** (6): 769–789. Bibcode 1949PhRv...76..769F. doi:10.1103/PhysRev.76.769.
- Fowler, Michael (1999). *The Bohr Atom*. University of Virginia.
- Heisenberg, Werner (1958). *Physics and Philosophy*. Harper and Brothers. ISBN 0-06-130549-9. LCCN 99010404.
- Lakshmbala, S. (2004). "Heisenberg, Matrix Mechanics and the Uncertainty Principle". *Resonance, Journal of Science Education* **9** (8).
- Liboff, Richard L. (1992). *Introductory Quantum Mechanics* (2nd ed.).
- Lindsay, Robert Bruce; Margenau, Henry (1957). *Foundations of Physics*. Dover. ISBN 0-918024-18-8. LCCN 57014416.
- McEvoy, J. P.; Zarate, Oscar. *Introducing Quantum Theory*. ISBN 1-874166-37-4.
- Müller-Kirsten, H. J. W. (2012). *Introduction to Quantum Mechanics: Schrödinger Equation and Path Integral* (2nd ed.). World Scientific. ISBN 978-981-4397-74-2.
- Nave, Carl Rod (2005). "Quantum Physics" (<http://hyperphysics.phy-astr.gsu.edu/hbase/quacon.html#quacon>). *HyperPhysics*. Georgia State University.
- Peat, F. David (2002). *From Certainty to Uncertainty: The Story of Science and Ideas in the Twenty-First Century*. Joseph Henry Press.
- Reichenbach, Hans (1944). *Philosophic Foundations of Quantum Mechanics*. University of California Press. ISBN 0-486-40459-5. LCCN a44004471.
- Schlipp, Paul Arthur (1949). *Albert Einstein: Philosopher-Scientist*. Tudor Publishing Company. LCCN 50005340.
- *Scientific American Reader*, 1953.
- Sears, Francis Weston (1949). *Optics* (3rd ed.). Addison-Wesley. ISBN 0-19-504601-3. LCCN 51001018.
- Shimony, A. (1983). "(title not given in citation)". *Foundations of Quantum Mechanics in the Light of New Technology* (S. Kamefuchi et al., eds.). Tokyo: Japan Physical Society. pp. 225.; cited in: Popescu, Sandu; Daniel Rohrlich (1996). "Action and Passion at a Distance: An Essay in Honor of Professor Abner Shimony". arXiv:quant-ph/9605004 [quant-ph].

- Tavel, Morton; Tavel, Judith (illustrations) (2002). *Contemporary physics and the limits of knowledge* (<http://books.google.com/?id=SELS0HbIhjYC&pg=PA200&dq=Wave+function+collapse>). Rutgers University Press. ISBN 978-0-8135-3077-2.
- Van Vleck, J. H., 1928, "The Correspondence Principle in the Statistical Interpretation of Quantum Mechanics," *Proc. Nat. Acad. Sci.* 14: 179.
- Wheeler, John Archibald; Feynman, Richard P. (1949). "Classical Electrodynamics in Terms of Direct Interparticle Action". *Reviews of Modern Physics* **21** (3): 425–433. Bibcode 1949RvMP...21..425W. doi:10.1103/RevModPhys.21.425.
- Wieman, Carl; Perkins, Katherine (2005). "Transforming Physics Education". *Physics Today*.
- Westmoreland; Benjamin Schumacher (1998). "Quantum Entanglement and the Nonexistence of Superluminal Signals". arXiv:quant-ph/9801014 [quant-ph].
- Bronner, Patrick; Strunz, Andreas; Silberhorn, Christine; Meyn, Jan-Peter (2009). "Demonstrating quantum random with single photons". *European Journal of Physics* **30** (5): 1189–1200. Bibcode 2009EJPh...30.1189B. doi:10.1088/0143-0807/30/5/026.

## Further reading

The following titles, all by working physicists, attempt to communicate quantum theory to lay people, using a minimum of technical apparatus.

- Malin, Shimon (2012). *Nature Loves to Hide: Quantum Physics and the Nature of Reality, a Western Perspective* (Revised ed.). World Scientific. ISBN 978-981-4324-57-1.
- Jim Al-Khalili (2003) *Quantum: A Guide for the Perplexed*. Weidenfield & Nicholson.
- Richard Feynman (1985) *QED: The Strange Theory of Light and Matter*. Princeton University Press. ISBN 0-691-08388-6
- Ford, Kenneth (2005) *The Quantum World*. Harvard Univ. Press. Includes elementary particle physics.
- Ghirardi, GianCarlo (2004) *Sneaking a Look at God's Cards*, Gerald Malsbary, trans. Princeton Univ. Press. The most technical of the works cited here. Passages using algebra, trigonometry, and bra-ket notation can be passed over on a first reading.
- Tony Hey and Walters, Patrick (2003) *The New Quantum Universe*. Cambridge Univ. Press. Includes much about the technologies quantum theory has made possible.
- Vladimir G. Ivancevic, Tijana T. Ivancevic (2008) *Quantum leap: from Dirac and Feynman, across the universe, to human body and mind*. World Scientific Publishing Company. Provides an intuitive introduction in non-mathematical terms and an introduction in comparatively basic mathematical terms.
- N. David Mermin (1990) "Spooky actions at a distance: mysteries of the QT" in his *Boojums all the way through*. Cambridge Univ. Press: 110–176. The author is a rare physicist who tries to communicate to philosophers and humanists.
- Roland Omnes (1999) *Understanding Quantum Mechanics*. Princeton Univ. Press.
- Victor Stenger (2000) *Timeless Reality: Symmetry, Simplicity, and Multiple Universes*. Buffalo NY: Prometheus Books. Chpts. 5–8.
- Martinus Veltman (2003) *Facts and Mysteries in Elementary Particle Physics*. World Scientific Publishing Company.
- Brian Cox and Jeff Forshaw (2011) *The Quantum Universe*. Allen Lane.



## External links

- Takada, Kenjiro, Emeritus professor at Kyushu University, " Microscopic World – Introduction to Quantum Mechanics. ([http://www.kutl.kyushu-u.ac.jp/seminar/MicroWorld1\\_E/MicroWorld\\_1\\_E.html](http://www.kutl.kyushu-u.ac.jp/seminar/MicroWorld1_E/MicroWorld_1_E.html))"
- Quantum Theory. (<http://www.encyclopedia.com/doc/1E1-quantum.html>)
- Quantum Mechanics. (<http://www.aip.org/history/heisenberg/p07.htm>)
- The spooky quantum ([http://www.imamu.edu.sa/Scientific\\_selections/abstracts/Physics/THE\\_SPOOKY\\_QUANTUM.pdf](http://www.imamu.edu.sa/Scientific_selections/abstracts/Physics/THE_SPOOKY_QUANTUM.pdf))
- Everything you wanted to know about the quantum world. (<http://www.newscientist.com/channel/fundamentals/quantum-world>) From the *New Scientist*.
- This Quantum World. (<http://thisquantumworld.com/>)
- The Quantum Exchange (<http://www.compadre.org/quantum>) (tutorials and open source learning software).
- Atoms and the Periodic Table (<http://www.chem1.com/acad/webtext/atoms/>)
- Theoretical Physics wiki (<http://theoreticalphysics.wetpaint.com>)
- " Uncertainty Principle, (<http://www.thebigview.com/spacetime/index.html>)" a recording of Werner Heisenberg's voice.
- Single and double slit interference ([http://intro.phys.psu.edu/class/251Labs/10\\_Interference\\_&\\_Diffraction/Single\\_and\\_Double-Slit\\_Interference.pdf](http://intro.phys.psu.edu/class/251Labs/10_Interference_&_Diffraction/Single_and_Double-Slit_Interference.pdf))
- Time-Evolution of a Wavepacket in a Square Well (<http://demonstrations.wolfram.com/TimeEvolutionOfAWavepacketInASquareWell/>) An animated demonstration of a wave packet dispersion over time.
- Experiments with single photons (<http://www.didaktik.physik.uni-erlangen.de/quantumlab/english/>) An introduction into quantum physics with interactive experiments
- Hitachi video recording of double-slit experiment done with electrons (<http://www.youtube.com/watch?v=oxknfn97vFE>) - You can see the interference pattern build up over time.

# Quantum Zeno effect

---

The **quantum Zeno effect** is a name coined by George Sudarshan and Baidyanath Misra of the University of Texas in 1977 in their analysis of the situation in which an unstable particle, if observed continuously, will never decay.<sup>[1]</sup> One can "freeze" the evolution of the system by measuring it frequently enough in its (known) initial state. The meaning of the term has since expanded, leading to a more technical definition in which time evolution can be suppressed not only by measurement: the quantum Zeno effect is the suppression of unitary time evolution caused by quantum decoherence in quantum systems provided by a variety of sources: measurement, interactions with the environment, stochastic fields, and so on.<sup>[2]</sup> As an outgrowth of study of the quantum Zeno effect, it has become clear that applying a series of sufficiently strong and fast pulses with appropriate symmetry can also *decouple* a system from its decohering environment.<sup>[3]</sup>

The name comes from Zeno's arrow paradox which states that, since an arrow in flight is not seen to move during any single instant, it cannot possibly be moving at all.<sup>[4]</sup>

An earlier theoretical exploration of this effect of measurement was published in 1974 by Degasperis *et al.*<sup>[5]</sup> and Alan Turing described it in 1954:<sup>[6]</sup>

It is easy to show using standard theory that if a system starts in an eigenstate of some observable, and measurements are made of that observable  $N$  times a second, then, even if the state is not a stationary one, the probability that the system will be in the same state after, say, one second, tends to one as  $N$  tends to infinity; that is, that continual observations will prevent motion ...

— Alan Turing as quoted by A. Hodges in *Alan Turing: Life and Legacy of a Great Thinker* p. 54

resulting in the earlier name **Turing paradox**. The idea is contained in the early work by John von Neumann, sometimes called the *reduction postulate*.<sup>[7]</sup>

According to the reduction postulate, each measurement causes the wavefunction to "collapse" to a pure eigenstate of the measurement basis. In the context of this effect, an "observation" can simply be the absorption of a particle, without an observer in any conventional sense. However, there is controversy over the interpretation of the effect, sometimes referred to as the "measurement problem" in traversing the interface between microscopic and macroscopic.<sup>[8][9]</sup>

Closely related (and sometimes not distinguished from the quantum Zeno effect) is the **watchdog effect**, in which the time evolution of a system is affected by its continuous coupling to the environment.<sup>[10][11]</sup>

## Description

Unstable quantum systems are predicted to exhibit a short time deviation from the exponential decay law.<sup>[12][13]</sup> This universal phenomenon has led to the prediction that frequent measurements during this nonexponential period could inhibit decay of the system, one form of the **quantum Zeno effect**. Subsequently, it was predicted that an *enhancement* of decay due to frequent measurements could be observed under somewhat more general conditions, leading to the so-called **anti-Zeno effect**.<sup>[14]</sup>

In quantum mechanics, the interaction mentioned is called "measurement" because its result can be interpreted in terms of classical mechanics. Frequent measurement prohibits the transition. It can be a transition of a particle from one half-space to another (which could be used for atomic mirror in an atomic nanoscope<sup>[15]</sup>) as in the time of arrival problem<sup>[16]</sup>,<sup>[17]</sup> a transition of a photon in a waveguide from one mode to another, and it can be a transition of an atom from one quantum state to another. It can be a transition from the subspace without decoherent loss of a q-bit to a state with a q-bit lost in a quantum computer.<sup>[18][19]</sup> In this sense, for the q-bit correction, it is sufficient to determine whether the decoherence has already occurred or not. All these can be considered as applications of the Zeno effect.<sup>[20]</sup> By its nature, the effect appears only in systems with distinguishable quantum states, and hence is inapplicable to classical phenomena and macroscopic bodies.

---

## Various realizations and general definition

The treatment of the Zeno effect as a paradox is not limited to the processes of quantum decay. In general, the term **Zeno effect** is applied to various transitions, and sometimes these transitions may be very different from a mere "decay" (whether exponential or non-exponential).

One realization refers to the observation of an object (Zeno's arrow, or any quantum particle) as it leaves some region of space. In the 20th century, the trapping (confinement) of a particle in some region by its observation outside the region was considered as nonsensical, indicating some non-completeness of quantum mechanics.<sup>[21]</sup> Even as late as 2001, confinement by absorption was considered as a paradox.<sup>[22]</sup> Later, similar effects of the suppression of Raman scattering was considered an expected *effect*,<sup>[23][24][25]</sup> not a paradox at all. The absorption of a photon at some wavelength, the release of a photon (for example one that has escaped from some mode of a fiber), or even the relaxation of a particle as it enters some region, are all processes that can be interpreted as measurement. Such a measurement suppresses the transition, and is called the Zeno effect in the scientific literature.

In order to cover all of these phenomena (including the original effect of suppression of quantum decay), the Zeno effect can be defined as a class of phenomena in which some transition is suppressed by an interaction — one that allows the interpretation of the resulting state in the terms *transition did not yet happen* and *transition has already occurred*, or *The proposition that the evolution of a quantum system is halted* if the state of the system is continuously measured by a macroscopic device to check whether the system is still in its initial state.<sup>[26]</sup>

## Periodic measurement of a quantum system

Consider a system in a state A, which is the eigenstate of some measurement operator. Say the system under free time evolution will decay with a certain probability into state B. If measurements are made periodically, with some finite interval between each one, at each measurement, the wave function collapses to an eigenstate of the measurement operator. Between the measurements, the system evolves away from this eigenstate into a superposition state of the states A and B. When the superposition state is measured, it will again collapse, either back into state A as in the first measurement, or away into state B. However, its probability of collapsing into state B, after a very short amount of time  $t$ , is proportional to  $t^2$ , since probabilities are proportional to squared amplitudes, and amplitudes behave linearly. Thus, in the limit of a large number of short intervals, with a measurement at the end of every interval, the probability of making the transition to B goes to zero.

According to decoherence theory, the collapse of the wave function is not a discrete, instantaneous event. A "measurement" is equivalent to strongly coupling the quantum system to the noisy thermal environment for a brief period of time, and continuous strong coupling is equivalent to frequent "measurement". The time it takes for the wave function to "collapse" is related to the decoherence time of the system when coupled to the environment. The stronger the coupling is, and the shorter the decoherence time, the faster it will collapse. So in the decoherence picture, a perfect implementation of the quantum Zeno effect corresponds to the limit where a quantum system is continuously coupled to the environment, and where that coupling is infinitely strong, and where the "environment" is an infinitely large source of thermal randomness.

## Experiments and discussion

Experimentally, strong suppression of the evolution of a quantum system due to environmental coupling has been observed in a number of microscopic systems.

In 1989, David J. Wineland and his group at NIST<sup>[27]</sup> observed the quantum Zeno effect for a two-level atomic system that is interrogated during its evolution. Approximately 5000  ${}^9\text{Be}^+$  ions were stored in a cylindrical Penning trap and laser cooled to below 250 mK. A resonant RF pulse was applied which, if applied alone, would cause the entire ground state population to migrate into an excited state. After the pulse was applied, the ions were monitored for photons emitted due to relaxation. The ion trap was then regularly "measured" by applying a sequence of

ultraviolet pulses, during the RF pulse. As expected, the ultraviolet pulses suppressed the evolution of the system into the excited state. The results were in good agreement with theoretical models. A recent review describes subsequent work in this arena.<sup>[28]</sup>

In 2001, Mark G. Raizen and his group at the University of Texas at Austin, observed the quantum Zeno and anti-Zeno effects for an unstable quantum system,<sup>[29]</sup> as originally proposed by Sudarshan and Misra.<sup>[1]</sup> Ultracold sodium atoms were trapped in an accelerating optical lattice and the loss due to tunneling was measured. The evolution was interrupted by reducing the acceleration, thereby stopping quantum tunneling. The group observed suppression or enhancement of the decay rate, depending on the regime of measurement.

The Quantum Zeno Effect is used in commercial atomic magnetometers and naturally by birds' magnetic compass sensory mechanism (magnetoreception).<sup>[30]</sup>

It is still an open question how closely one can approach the limit of an infinite number of interrogations due to the Heisenberg uncertainty involved in shorter measurement times. In 2006, Streed *et al.* at MIT observed the dependence of the Zeno effect on measurement pulse characteristics.<sup>[31]</sup>

The interpretation of experiments in terms of the "Zeno effect" helps describe the origin of a phenomenon. Nevertheless, such an interpretation does not bring any principally new features not described with the Schrödinger equation of the quantum system.<sup>[32][33]</sup>

Even more, the detailed description of experiments with the "Zeno effect", especially at the limit of high frequency of measurements (high efficiency of suppression of transition, or high reflectivity of a ridged mirror) usually do not behave as expected for an idealized measurement,<sup>[15]</sup> and require analysis of the mechanism of the interaction.<sup>[34]</sup>

It was shown that the Quantum Zeno effect persists in the many-worlds and relative states interpretations of quantum mechanics.<sup>[35]</sup>

## Significance to cognitive science

The quantum Zeno effect (with its own controversies related to the problem of measurement) is becoming a central concept in the exploration of controversial theories of quantum mind consciousness within the discipline of cognitive science. In his book *Mindful Universe* (2007), Henry Stapp claims that the mind holds the brain in a superposition of states using the quantum Zeno effect. He advances that this phenomenon is the principal method by which the conscious can effect change, a possible solution to the mind-body dichotomy. Stapp and co-workers do not claim finality of their theory, but only:<sup>[36]</sup>

The new framework, unlike its classic-physics-based predecessor, is erected directly upon, and is compatible with, the prevailing principles of physics.

Needless to say, such conjectures have their opponents, serving perhaps to create more furor, rather than less, for example, see Bourget.<sup>[37]</sup> Recent work<sup>[38]</sup> criticizes Stapp's model in two aspects: (1) The mind in Stapp's model does not have its own wavefunction or density matrix, but nevertheless can act upon the brain using projection operators. Such usage is not compatible with standard quantum mechanics because one can attach any number of ghostly minds to any point in space that act upon physical quantum systems with any projection operators. Therefore Stapp's model does not build upon "the prevailing principles of physics", but negates them.<sup>[38]</sup> (2) Stapp's claim that quantum Zeno effect is robust against environmental decoherence directly contradicts a basic theorem in quantum information theory according to which acting with projection operators upon the density matrix of a quantum system can never decrease the Von Neumann entropy of the system, but can only increase it.<sup>[38]</sup> Indeed, already in 1993 it was shown by M. J. Gagen and colleagues<sup>[39]</sup> that the quantum Zeno effect is easily destroyed by noise and that a two-level system becomes a "random telegraph", i.e. the evolution of the system is not suppressed as required for quantum Zeno effect, instead the system jumps randomly between the two states.

A summary of the situation is provided by Davies:<sup>[40]</sup>

There have been many claims that quantum mechanics plays a key role in the origin and/or operation of biological organisms, beyond merely providing the basis for the shapes and sizes of biological molecules and their chemical affinities...The case for quantum biology remains one of "not proven." There are many suggestive experiments and lines of argument indicating that some biological functions operate close to, or within, the quantum regime, but as yet no clear-cut example has been presented of non-trivial quantum effects at work in a key biological process.

The significance of the Zeno effect in determining the rate of quantum decoherence in biological systems remains unknown.

## External links

- Zeno.qcl<sup>[41]</sup> A computer program written in QCL which demonstrates the Quantum Zeno effect

## References

- [1] Sudarshan, E.C.G.; Misra, B. (1977). "The Zeno's paradox in quantum theory" (<http://scitation.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=JMAPAQ000018000004000756000001&idtype=cvips&gifs=yes>). *Journal of Mathematical Physics* **18** (4): 756–763. Bibcode 1977JMP...18..756M. doi:10.1063/1.523304.
- [2] T. Nakanishi, K. Yamane, and M. Kitano: *Absorption-free optical control of spin systems: the quantum Zeno effect in optical pumping* Phys. Rev. A **65**, 013404 (2001) (<http://arxiv.org/abs/quant-ph/0103034v2>)
- [3] P. Facchi, D. A. Lidar, & S. Pascazio *Unification of dynamical decoupling and the quantum Zeno effect* Physical Review A **69**, 032314 (2004) (<http://qserver.usc.edu/group/wp-content/uploads/2007/02/pr690323142004.pdf>)
- [4] The idea depends on the *instant of time*, a kind of freeze-motion idea that the arrow is "strobed" at each instant and is seemingly stationary, so how can it move in a succession of stationary events?
- [5] A. Degasperis, L. Fonda & G.C. Ghirardi (1974). "Does the lifetime of an unstable system depend on the measuring apparatus?" (<http://www.springerlink.com/content/4064h6xv57162038/>). *Il Nuovo Cimento A* **21** (3): 471–484. Bibcode 1974NCimA...21..471D. doi:10.1007/BF02731351.
- [6] Christof Teuscher & Douglas Hofstadter (2004). *Alan Turing: Life and Legacy of a Great Thinker* ([http://books.google.com/?id=th0\\_ipQKMGMC&pg=PA54&dq=Turing+paradox](http://books.google.com/?id=th0_ipQKMGMC&pg=PA54&dq=Turing+paradox)). Springer. p. 54. ISBN 3-540-20020-7. .
- [7] John von Neumann (1932). *Mathematische Grundlagen der Quantenmechanik*. Springer. Chapter V.2. ISBN 3-540-59207-5. (See also: J. von Neumann (1955). *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press. p. 366. ISBN 0-691-02893-1.); Michael B. Mensky (2000). *Quantum Measurements and Decoherence* (<http://books.google.com/?id=Bo7jujIMqL8C&pg=PA80&dq=von+Neumann+Zeno+effect>). Springer. p. §4.1.1, pp. 315 ff. ISBN 0-7923-6227-6. .; Ch. Wunderlich & Ch. Balzer (Benjamin Bederson & H Walther, editors) (2003). *Advances in Atomic, Molecular, and Optical Physics: Vol. 49* (<http://books.google.com/?id=mmhJ37o8fdwC&pg=PA315&dq=von+Neumann+Zeno+effect>). Academic Press. ISBN 0-12-003849-8. .
- [8] George Greenstein & Arthur Zajonc (2005). *The Quantum Challenge: Modern Research on the Foundations of Quantum Mechanics* (<http://books.google.com/?id=5t0tm0FB1CsC&pg=PA231&dq=quantum+Zeno>). Jones & Bartlett Publishers. p. 237. ISBN 0-7637-2470-X. .
- [9] P. Facchi and S. Pascazio (2002). "Quantum Zeno subspaces". *Physical Review Letters* **89** (8). arXiv:quant-ph/0201115. Bibcode 2002PhRvL..89h0401F. doi:10.1103/PhysRevLett.89.080401
- [10] Partha Ghose (1999). *Testing Quantum Mechanics on New Ground* (<http://books.google.com/?id=GqRQYEPZRYwC&pg=PA114&dq=watchdog+effect>). Cambridge University Press. p. 107. ISBN 0-521-02659-8. .
- [11] Gennaro Auletta (2000). *Foundations and Interpretation of Quantum Mechanics* (<http://books.google.com/?id=ISAFY0LEKBM&pg=RA1-PA341&dq=watchdog+effect>). World Scientific. p. 341. ISBN 981-02-4614-5. .
- [12] Khalifin, L.A. (1958). "Contribution to the decay theory of a quasi-stationary state" ([http://www.osti.gov/energycitations/product.biblio.jsp?osti\\_id=4318804](http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=4318804)). *Soviet Phys. JETP* **6**: 1053. Bibcode 1958JETP...6.1053K.
- [13] Steven R. Wilkinson, Cyrus F. Bharucha, Martin C. Fischer, Kirk W. Madison, Patrick R. Morrow, Qian Niu, Bala Sundaram & Mark G. Raizen (1997). "Experimental evidence for non-exponential decay in quantum tunnelling" ([http://www.phytem.ens-cachan.fr/telechargement/Optique\\_Quantique/Raizen\\_decay.pdf](http://www.phytem.ens-cachan.fr/telechargement/Optique_Quantique/Raizen_decay.pdf)). *Nature* **387** (6633): 575 ff. Bibcode 1997Natur.387..575W. doi:10.1038/42418.
- [14] This definition is a paraphrase of the introductory remarks in Raizen *et al.*'s paper cited later in this article.
- [15] D.Kouznetsov; H. Oberst, K. Shimizu, A. Neumann, Y. Kuznetsova, J.-F. Bisson, K. Ueda, S. R. J. Brueck (2006). "Ridged atomic mirrors and atomic nanoscope" (<http://stacks.iop.org/0953-4075/39/1605>). *JOPB* **39** (7): 1605–1623. Bibcode 2006JPhB...39.1605K. doi:10.1088/0953-4075/39/7/005. .
- [16] Allcock, J. (1969). "The time of arrival in quantum mechanics I. Formal considerations" (<http://www.sciencedirect.com/science/article/pii/0003491669902516>). *Ann. Phys. (N.Y.)* **53**: 253–285. Bibcode 1969AnPhy..53..253A. doi:10.1016/0003-4916(69)90251-6. .
- [17] Echanobe, J.; del Campo, A.; Muga, J. G. (2008). "Disclosing hidden information in the quantum Zeno effect: Pulsed measurement of the quantum time of arrival" (<http://link.aps.org/doi/10.1103/PhysRevA.77.032112>). *Physical Review A* **77**: 032112. arXiv:0712.0670.

- Bibcode 2008PhRvA..77c2112E. doi:10.1103/PhysRevA.77.032112. .
- [18] Joachim Stolze & Dieter Suter (2008). *Quantum computing: a short course from theory to experiment* (<http://books.google.com/?id=VkpGN1z15bcC&printsec=frontcover&dq=intitle:Quantum+intitle:Computing+inauthor:Stolze>) (2nd ed.). Wiley-VCH. p. §7.3.7 p.99. ISBN 3-527-40787-1. .
- [19] Quantum computer: URL: <http://www.physorg.com/news11087.html>
- [20] JD Franson; B. C. Jacobs, and T. B. Pittman (2006). "Quantum computing using single photons and the Zeno effect" ([http://userpages.umbc.edu/~pittmtb1/pdf\\_publications/PRA\\_Zeno\\_Gates.pdf](http://userpages.umbc.edu/~pittmtb1/pdf_publications/PRA_Zeno_Gates.pdf)). *PRA* **70** (6): 062302. arXiv:quant-ph/0408097. Bibcode 2004PhRvA..70f2302F. doi:10.1103/PhysRevA.70.062302. .
- [21] B.Mielnik (1994). "The screen problem". *Foundations of physics* **24** (8): 1113–1129. Bibcode 1994FoPh...24.1113M. doi:10.1007/BF02057859. .
- [22] K.Yamane, M.Ito, M.Kitano (2001). "Quantum Zeno effect in optical fibers". *Optics Communications* **192** (3–6): 299–307. Bibcode 2001OptCo.192..299Y. doi:10.1016/S0030-4018(01)01192-0. .
- [23] K.Thun, J.Perina, J.Krepelka. (2002). "Quantum Zeno effect in Raman scattering". *Physics Letters A* **299** (1): 19–30. Bibcode 2002PhLA..299...19T. doi:10.1016/S0375-9601(02)00629-1. .
- [24] Peina, J (2004). "Quantum Zeno effect in cascaded parametric down-conversion with losses". *Physics Letters A* **325** (1): 16–20. Bibcode 2004PhLA..325...16P. doi:10.1016/j.physleta.2004.03.026. .
- [25] D.Kouznetsov; H.Oberst (2005). "Reflection of Waves from a Ridged Surface and the Zeno Effect" (<http://annex.jsap.or.jp/OSJ/opticalreview/TOC-Lists/vol12/12e0363tx.htm>). *Optical Review* **12** (5): 1605–1623. Bibcode 2005OptRv..12..363K. doi:10.1007/s10043-005-0363-9. .
- [26] A.D.Panov (2001). "Quantum Zeno effect in spontaneous decay with distant detector". *Physics Letters A* **281**: 9. arXiv:quant-ph/0101031. Bibcode 2001PhLA..281....9P. doi:10.1016/S0375-9601(01)00094-9. .
- [27] W.M.Itano; D.J.Heinsen, J.J.Bokkinger, D.J.Wineland (1990). "Quantum Zeno effect" (<http://www.boulder.nist.gov/timefreq/general/pdf/858.pdf>). *PRA* **41** (5): 2295–2300. Bibcode 1990PhRvA..41.2295I. doi:10.1103/PhysRevA.41.2295. .
- [28] D. Leibfried, R. Blatt, C. Monroe & D. Wineland: *Quantum dynamics of single trapped ions* Reviews of Modern Physics, Vol. 75, January 2003 p. 281 ff ([http://cua.mit.edu/8.422\\_old/PHYSICS-leibfried-blatt-monroe-wineland-quantum-dynamics-of-single-trapped-ions-rmp75-p281-2003.pdf](http://cua.mit.edu/8.422_old/PHYSICS-leibfried-blatt-monroe-wineland-quantum-dynamics-of-single-trapped-ions-rmp75-p281-2003.pdf))
- [29] M. C. Fischer, B. Gutiérrez-Medina, and M. G. Raizen *Observation of the Quantum Zeno and Anti-Zeno Effects in an Unstable System* Phys. Rev. Lett. 87, 040402 (2001) (<http://arxiv.org/abs/quant-ph/0104035v1>)
- [30] *Quantum Zeno Effect Underpinning the Radical-Ion-Pair Mechanism of Avian Magnetoreception* by Iannis K. Kominis (<http://arxiv.org/abs/0804.2646>)
- [31] Erik W. Streed, Jongchul Mun, Micah Boyd, Gretchen K. Campbell, Patrick Medley, Wolfgang Ketterle & David E. Pritchard: *Continuous and Pulsed Quantum Zeno Effect* Physical Phys. Rev. Lett. 97, 260402 (2006) (<http://arxiv.org/abs/cond-mat/0606430v1>)
- [32] T. Petrosky, S. Tasaki and I. Prigogine: *Quantum Zeno effect* Physics Letters A Vol. 151, 109 (1990) (<http://www.sciencedirect.com/science/article/pii/037596019090173L>)
- [33] T. Petrosky, S. Tasaki and I. Prigogine: *Quantum Zeno effect* Physica A Vol. 170, 306 (1991) (<http://www.sciencedirect.com/science/article/pii/037843719190048H>)
- [34] K. Koshino and A Shimizu (2005) *Quantum Zeno effect by general measurements* Physics Reports, Volume 412, Issue 4, p. 191-275 (<http://arxiv.org/abs/quant-ph/0411145v2>)
- [35] D. Home; M. A. B. Whitaker (1987). "The many-worlds and relative states interpretations of quantum mechanics, and the quantum Zeno paradox" (<http://www.iop.org/EJ/abstract/0305-4470/20/11/036>). *Journal of Physics A* **20** (11): 3339–3345. Bibcode 1987JPhA...20.3339H. doi:10.1088/0305-4470/20/11/036. .
- [36] J.M. Schwartz, H.P. Stapp & M. Beauregard (2005) "Quantum physics in neuroscience and psychology: a neurophysical model of mind-brain interaction", *Philosophical Transactions of the Royal Society of London B* Volume 360, Number 1458 / June 29, 2005 Pages 1309-1327 DOI 10.1098/rstb.2004.1598 (<http://journals.royalsociety.org/content/fl8g2b86ykbxht3m/>)
- [37] Bourget, David: *Quantum Leaps in Philosophy of Mind: A Critique of Stapp's Theory* Journal of Consciousness Studies, Volume 11, Number 12, 2004 , pp. 17-42(26) (<http://www.newdualism.org/papers/D.Bourget/QLPM.htm>)
- [38] Georgiev D (2012). "Mind efforts, quantum Zeno effect and environmental decoherence" (<http://www.neuroquantology.com/index.php/journal/article/view/552>). *NeuroQuantology* **10** (3): 374–388. .
- [39] Gagen MJ, Wiseman HM, Milburn GJ (1993). "Continuous position measurements and the quantum Zeno effect" (<http://link.aps.org/doi/10.1103/PhysRevA.48.132>). *Physical Review A* **48** (1): 132–142. doi:10.1103/PhysRevA.48.132. PMID 9909581. .
- [40] PCW Davies: *Does quantum mechanics play a non-trivial role in life?* BioSystems 78 (2004) 69–79 ([http://aca.mq.edu.au/PaulDavies/publications/papers/Does quantum mechanics play a non trivial role in life' BioSystems paper.pdf](http://aca.mq.edu.au/PaulDavies/publications/papers/Does%20quantum%20mechanics%20play%20a%20non%20trivial%20role%20in%20life%20BioSystems%20paper.pdf))
- [41] <http://www.lysator.liu.se/~jc/zeno.qcl.html>
- Quantum Zeno effect on arxiv.org (<http://xstructure.inr.ac.ru/x-bin/theme3.py?level=1&index1=-209310>)

# Observer effect (physics)

---

In science, the term **observer effect** refers to changes that the act of observation will make on a phenomenon being observed. This is often the result of instruments that, by necessity, alter the state of what they measure in some manner. A commonplace example is checking the pressure in an automobile tire; this is difficult to do without letting out some of the air, thus changing the pressure. This effect can be observed in many domains of physics.

The observer effect on a physical process can often be reduced to insignificance by using better instruments or observation techniques. However in quantum mechanics, which deals with very small objects, it is not possible to observe a system without changing the system, so the observer must be considered part of the system being observed.

## Particle physics

For an electron to become detectable, a photon must first interact with it, and this interaction will change the path of that electron. It is also possible for other, less direct means of measurement to affect the electron.

## Electronics

In electronics, ammeters and voltmeters are usually wired in series or parallel to the circuit, and so by their very presence affect the current or the voltage they are measuring by way of presenting an additional real or complex load to the circuit, thus changing the transfer function and behaviour of the circuit itself. Even a more passive device such as a current clamp, which measures the wire current without coming into physical contact with the wire, affects the current through the circuit being measured because the inductance is mutual.

## Thermodynamics

In thermodynamics, a standard mercury-in-glass thermometer must absorb or give up some thermal energy to record a temperature, and therefore changes the temperature of the body which it is measuring.

## Quantum mechanics

The theoretical foundation of the concept of measurement in quantum mechanics is a contentious issue deeply connected to the many interpretations of quantum mechanics. A key topic is that of wave function collapse, for which some interpretations assert that measurement causes a discontinuous change into a non-quantum state, which no longer evolves. The superposition principle ( $\psi = \sum_n a_n \psi_n$ ) of quantum physics says that for a wave function  $\psi$ , a measurement will give a state of the quantum system of one of the  $m$  possible eigenvalues  $f_n$ ,  $n=1,2,\dots,m$ , of the operator  $\hat{F}$  which is part of the eigenfunctions  $\psi_n$ ,  $n=1,2,\dots,n$ . Once we have measured the system, we know its current state and this stops it from being in one of its other states.<sup>[1]</sup> This means that the type of measurement that we do on the system affects the end state of the system. An experimentally studied situation related to this is the quantum Zeno effect, in which a quantum state that would decay if left alone but does not decay because of its continuous observation. The dynamics of a quantum system under continuous observation is described by a quantum stochastic master equation known as the Belavkin equation.<sup>[2] [3] [4]</sup>

A consequence of Bell's theorem is that measurement on one of two entangled particles can appear to have a nonlocal effect on the opposite particle. Additional problems related to decoherence arise when the observer too is modeled as a quantum system.

The uncertainty principle has been frequently confused with the observer effect, evidently even by its originator, Werner Heisenberg.<sup>[5]</sup> The uncertainty principle in its standard form actually describes how precisely we may measure the position and momentum of a particle at the same time — if we increase the precision in measuring one

---



quantity, we are forced to lose precision in measuring the other.<sup>[6]</sup> An alternative version of the uncertainty principle,<sup>[7]</sup> more in the spirit of an observer effect,<sup>[8]</sup> fully accounts for the disturbance the observer has on a system and the error incurred, although this is not how the term "uncertainty principle" is most commonly used in practice.

## References

- [1] B.D'Espagnat, P.Eberhard, W.Schommers, F.Selleri. *Quantum Theory and Pictures of Reality*. Springer-Verlag, 1989, ISBN 3-540-50152-5
  - [2] V. P. Belavkin (1989). "A new wave equation for a continuous non-demolition measurement". *Physics Letters A* **140** (7-8): 355–358. arXiv:quant-ph/0512136. doi:10.1016/0375-9601(89)90066-2.
  - [3] Howard J. Carmichael (1993). *An Open Systems Approach to Quantum Optics*. Springer-Verlag.
  - [4] Michel Bauer, Denis Bernard, Tristan Benoist. *Iterated Stochastic Measurements*. arXiv:1210.0425.
  - [5] Furuta, Aya (2012), "One Thing Is Certain: Heisenberg's Uncertainty Principle Is Not Dead" (<http://www.scientificamerican.com/article/cfm?id=heisenbergs-uncertainty-principle-is-not-dead>), *Scientific American*,
  - [6] Heisenberg, W. (1930), *Physikalische Prinzipien der Quantentheorie*, Leipzig: Hirzel English translation *The Physical Principles of Quantum Theory*. Chicago: University of Chicago Press, 1930.
  - [7] Ozawa, Masanao (2003), "Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement", *Physical Review A* **67**, arXiv:quant-ph/0207121, Bibcode 2003PhRvA..67d2105O, doi:10.1103/PhysRevA.67.042105
  - [8] V. P. Belavkin (1992). "Quantum continual measurements and a posteriori collapse on CCR". *Communications in Mathematical Physics* **146** (3): 611–635. arXiv:math-ph/0512070. doi:10.1007/BF02097018.
- The observer effect by Massimiliano Sassoli de Bianchi <http://arxiv.org/abs/1109.3536>

# Uncertainty principle

In quantum mechanics, the **uncertainty principle** is any of a variety of mathematical inequalities asserting a fundamental limit to the precision with which certain pairs of physical properties of a particle, such as position  $x$  and momentum  $p$ , can be known simultaneously. The more precisely the position of some particle is determined, the less precisely its momentum can be known, and vice versa.<sup>[1]</sup> The original heuristic argument that such a limit should exist was given by Werner Heisenberg in 1927, after whom it is sometimes named the **Heisenberg principle**. A more formal inequality relating the standard deviation of position  $\sigma_x$  and the standard deviation of momentum  $\sigma_p$  was derived by Earle Hesse Kennard<sup>[2]</sup> later that year (and independently by Hermann Weyl<sup>[3]</sup> in 1928),

$$\sigma_x \sigma_p \geq \frac{\hbar}{2},$$

where  $\hbar$  is the reduced Planck constant.

Historically, the uncertainty principle has been confused<sup>[4][5]</sup> with a somewhat similar effect in physics, called the observer effect, which notes that measurements of certain systems cannot be made without affecting the systems. Heisenberg offered such an observer effect at the quantum level (see below) as a physical "explanation" of quantum uncertainty.<sup>[6]</sup> It has since become clear, however, that the uncertainty principle is inherent in the properties of all wave-like systems, and that it arises in quantum mechanics simply due to the matter wave nature of all quantum objects. Thus, *the uncertainty principle actually states a fundamental property of quantum systems, and is not a statement about the observational success of current technology.*<sup>[7]</sup> It must be emphasized that *measurement* does not mean only a process in which a physicist-observer takes part, but rather any interaction between classical and quantum objects regardless of any observer.<sup>[8]</sup>

Since the uncertainty principle is such a basic result in quantum mechanics, typical experiments in quantum mechanics routinely observe aspects of it. Certain experiments, however, may deliberately test a particular form of the uncertainty principle as part of their main research program. These include, for example, tests of number-phase uncertainty relations in superconducting<sup>[9]</sup> or quantum optics<sup>[10]</sup> systems. Applications are for developing extremely low noise technology such as that required in gravitational-wave interferometers.<sup>[11]</sup>

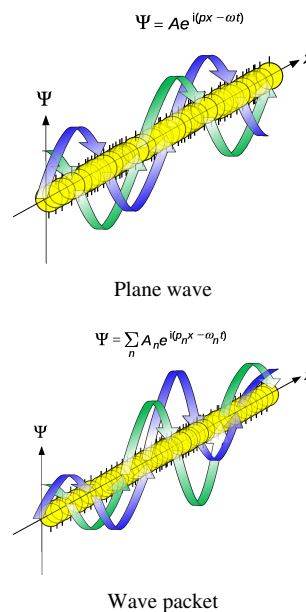
## Introduction

The uncertainty principle can be interpreted in either the wave mechanics or matrix mechanics formalisms of quantum mechanics. Although the principle is more visually intuitive in the wave mechanics formalism, it was first derived and is more easily generalized in the matrix mechanics formalism. We will attempt to motivate the principle in the two frameworks.

Mathematically, the uncertainty relation between position and momentum arises because the expressions of the wavefunction in the two corresponding bases are Fourier transforms of one another (i.e., position and momentum are conjugate variables). A similar tradeoff between the variances of Fourier conjugates arises wherever Fourier analysis is needed, for example in sound waves. A pure tone is a sharp spike at a single frequency. Its Fourier transform gives the shape of the sound wave in the time domain, which is a completely delocalized sine wave. In quantum mechanics, the two key points are that the position of the particle takes the form of a matter wave, and momentum is its Fourier conjugate, assured by the de Broglie relation  $p = \hbar k$ , where  $k$  is the wavenumber.

In the mathematical formulation of quantum mechanics, any pair of non-commuting self-adjoint operators representing observables are subject to similar uncertainty limits. An eigenstate of an observable represents the state of the wavefunction for a certain measurement value (the eigenvalue). For example, if a measurement of an observable  $A$  is taken then the system is in a particular eigenstate  $\Phi$  of that observable. The particular eigenstate of the observable  $A$  may not be an eigenstate of another observable  $B$ . If this is so, then it does not have a single associated measurement as the system is not in an eigenstate of the observable.<sup>[12]</sup>

## Wave mechanics interpretation



According to the de Broglie hypothesis, every object in the universe is a wave, a situation which gives rise to this phenomenon. The position of the particle is described by a wave function  $\Psi(x, t)$ . The time-independent wave function of a single-moded plane wave of wavenumber  $k_0$  or momentum  $p_0$  is

$$\psi(x) \propto e^{ik_0 x} = e^{ip_0 x/\hbar}$$

The Born rule states that this should be interpreted as a probability density function in the sense that the probability of finding the particle between  $a$  and  $b$  is

$$P[a \leq X \leq b] = \int_a^b |\psi(x)|^2 dx .$$

In the case of the single-moded plane wave,  $|\psi(x)|^2$  is a uniform distribution. In other words, the particle position is extremely uncertain in the sense that it could be essentially anywhere along the wave packet. Consider a wave function that is a sum of many waves, however, we may write this as

$$\psi(x) \propto \sum_n A_n e^{ip_n x/\hbar},$$

where  $A_n$  represents the relative contribution of the mode  $p_n$  to the overall total. The figures to the right show how with the addition of many plane waves, the wave packet can become more localized. We may take this a step further to the continuum limit, where the wave function is an integral over all possible modes

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \phi(p) \cdot e^{ipx/\hbar} dp,$$

with  $\phi(p)$  representing the amplitude of these modes and is called the wave function in momentum space. In mathematical terms, we say that  $\phi(p)$  is the *Fourier transform* of  $\psi(x)$  and that  $x$  and  $p$  are conjugate variables. Adding together all of these plane waves comes at a cost, namely the momentum has become less precise, having become a mixture of waves of many different momenta.

One way to quantify the precision of the position and momentum is the standard deviation  $\sigma$ . Since  $|\psi(x)|^2$  is a probability density function for position, we calculate its standard deviation.

We improved the precision of the position, i.e. reduced  $\sigma_x$ , by using many plane waves, thereby weakening the precision of the momentum, i.e. increased  $\sigma_p$ . Another way of stating this is that  $\sigma_x$  and  $\sigma_p$  have an inverse relationship or are at least bounded from below. This is the uncertainty principle, the exact limit of which, is the Kennard bound. Click the *show* button below to see a semi-formal derivation of the Kennard inequality using wave mechanics.

Proof of the Kennard inequality using wave mechanics

We are interested in the variances of position and momentum, defined as

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} x^2 \cdot |\psi(x)|^2 dx - \left( \int_{-\infty}^{\infty} x \cdot |\psi(x)|^2 dx \right)^2 \\ \sigma_p^2 &= \int_{-\infty}^{\infty} p^2 \cdot |\phi(p)|^2 dp - \left( \int_{-\infty}^{\infty} p \cdot |\phi(p)|^2 dp \right)^2. \end{aligned}$$

Without loss of generality, we will assume that the means vanish, which just amounts to a shift of the origin of our coordinates. (A more general proof that does not make this assumption is given below.) This gives us the simpler form

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} x^2 \cdot |\psi(x)|^2 dx \\ \sigma_p^2 &= \int_{-\infty}^{\infty} p^2 \cdot |\phi(p)|^2 dp. \end{aligned}$$

The function  $f(x) = x \cdot \psi(x)$  can be interpreted as a vector in a function space. We can define an inner product for a pair of functions  $u(x)$  and  $v(x)$  in this vector space:

$$\langle u|v \rangle = \int_{-\infty}^{\infty} u^*(x) \cdot v(x) dx,$$

where the asterisk denotes the complex conjugate.

With this inner product defined, we note that the variance for position can be written as

$$\sigma_x^2 = \int_{-\infty}^{\infty} |f(x)|^2 dx = \langle f|f \rangle.$$

We can repeat this for momentum by interpreting the function  $\tilde{g}(p) = p \cdot \phi(p)$  as a vector, but we can also take advantage of the fact that  $\psi(x)$  and  $\phi(p)$  are Fourier transforms of each other. We evaluate the inverse Fourier transform through integration by parts:

$$\begin{aligned}
 g(x) &= \frac{1}{\sqrt{2\pi\hbar}} \cdot \int_{-\infty}^{\infty} \tilde{g}(p) \cdot e^{ipx/\hbar} dp \\
 &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} p \cdot \phi(p) \cdot e^{ipx/\hbar} dp \\
 &= \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \left[ p \cdot \int_{-\infty}^{\infty} \psi(x) e^{-ipx/\hbar} dx \right] \cdot e^{ipx/\hbar} dp \\
 &= \frac{i}{2\pi} \int_{-\infty}^{\infty} \left[ \cancel{\psi(x) e^{-ipx/\hbar}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{d\psi(x)}{dx} e^{-ipx/\hbar} dx \right] \cdot e^{ipx/\hbar} dp \\
 &= \frac{-i}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{d\psi(x)}{dx} e^{-ipx/\hbar} dx e^{ipx/\hbar} dp \\
 &= \left( -i\hbar \frac{d}{dx} \right) \cdot \psi(x),
 \end{aligned}$$

where the canceled term vanishes because the wave function vanishes at infinity. Often the term  $-i\hbar \frac{d}{dx}$  is called the momentum operator in position space. Applying Parseval's theorem, we see that the variance for momentum can be written as

$$\sigma_p^2 = \int_{-\infty}^{\infty} |\tilde{g}(p)|^2 dp = \int_{-\infty}^{\infty} |g(x)|^2 dx = \langle g|g \rangle.$$

The Cauchy–Schwarz inequality asserts that

$$\sigma_x^2 \sigma_p^2 = \langle f|f \rangle \cdot \langle g|g \rangle \geq |\langle f|g \rangle|^2$$

The modulus squared of any complex number  $z$  can be expressed as

$$|z|^2 = \left( \text{Re}(z) \right)^2 + \left( \text{Im}(z) \right)^2 \geq \left( \text{Im}(z) \right)^2 = \left( \frac{z - z^*}{2i} \right)^2.$$

we let  $z = \langle f|g \rangle$  and  $z^* = \langle g|f \rangle$  and substitute these into the equation above to get

$$|\langle f|g \rangle|^2 \geq \left( \frac{\langle f|g \rangle - \langle g|f \rangle}{2i} \right)^2$$

All that remains is to evaluate these inner products.

$$\begin{aligned}
 \langle f|g \rangle - \langle g|f \rangle &= \int_{-\infty}^{\infty} \psi^*(x) x \cdot \left( -i\hbar \frac{d}{dx} \right) \psi(x) dx \\
 &\quad - \int_{-\infty}^{\infty} \psi^*(x) \left( -i\hbar \frac{d}{dx} \right) \cdot x \psi(x) dx \\
 &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \frac{d(x\psi(x))}{dx} \right] dx \\
 &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \left[ \left( -x \cdot \frac{d\psi(x)}{dx} \right) + \psi(x) + \left( x \cdot \frac{d\psi(x)}{dx} \right) \right] dx \\
 &= i\hbar \cdot \int_{-\infty}^{\infty} \psi^*(x) \psi(x) dx \\
 &= i\hbar \cdot \int_{-\infty}^{\infty} |\psi(x)|^2 dx \\
 &= i\hbar
 \end{aligned}$$

Plugging this into the above inequalities, we get

$$\sigma_x^2 \sigma_p^2 \geq |\langle f|g \rangle|^2 \geq \left( \frac{\langle f|g \rangle - \langle g|f \rangle}{2i} \right)^2 = \left( \frac{i\hbar}{2i} \right)^2 = \frac{\hbar^2}{4}$$

or taking the square root

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}.$$

Note that the only *physics* involved in this proof was that  $\psi(x)$  and  $\phi(p)$  are wave functions for position and momentum, which are Fourier transforms of each other. A similar result would hold for *any* pair of conjugate variables.

### Matrix mechanics interpretation

In matrix mechanics, observables such as position and momentum are represented by self-adjoint operators. When considering pairs of observables, one of the most important quantities is the *commutator*. For a pair of operators  $\hat{A}$  and  $\hat{B}$ , we may define their commutator as

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}.$$

In the case of position and momentum, the commutator is the canonical commutation relation

$$[\hat{x}, \hat{p}] = i\hbar.$$

The physical meaning of the non-commutativity can be understood by considering the effect of the commutator on position and momentum eigenstates. Let  $|\psi\rangle$  be a right eigenstate of position with a constant eigenvalue  $x_0$ . By definition, this means that  $\hat{x}|\psi\rangle = x_0|\psi\rangle$ . Applying the commutator to  $|\psi\rangle$  yields

$$[\hat{x}, \hat{p}]|\psi\rangle = (\hat{x}\hat{p} - \hat{p}\hat{x})|\psi\rangle = (\hat{x} - x_0\hat{I}) \cdot \hat{p}|\psi\rangle = i\hbar|\psi\rangle,$$

where  $\hat{I}$  is simply the identity operator. Suppose for the sake of proof by contradiction that  $|\psi\rangle$  is also a right eigenstate of momentum, with constant eigenvalue  $p_0$ . If this were true, then we could write

$$(\hat{x} - x_0\hat{I}) \cdot \hat{p}|\psi\rangle = (\hat{x} - x_0\hat{I}) \cdot p_0|\psi\rangle = (x_0\hat{I} - x_0\hat{I}) \cdot p_0|\psi\rangle = 0.$$

On the other hand, the canonical commutation relation requires that

$$[\hat{x}, \hat{p}]|\psi\rangle = i\hbar|\psi\rangle \neq 0.$$

This implies that no quantum state can be simultaneously both a position and a momentum eigenstate. When a state is measured, it is projected onto an eigenstate in the basis of the observable. For example, if a particle's position is measured, then the state exists at least momentarily in a position eigenstate. This means that the state is *not* in a momentum eigenstate, however, but rather exists as a sum of multiple momentum basis eigenstates. In other words the momentum must be less precise. The precision may be quantified by the standard deviations, defined by

$$\sigma_x = \sqrt{\langle \hat{x}^2 \rangle - \langle \hat{x} \rangle^2}$$

$$\sigma_p = \sqrt{\langle \hat{p}^2 \rangle - \langle \hat{p} \rangle^2}.$$

As with the wave mechanics interpretation above, we see a tradeoff between the precisions of the two, given by the uncertainty principle.

### Robertson–Schrödinger uncertainty relations

The most common general form of the uncertainty principle is the *Robertson uncertainty relation*.<sup>[13]</sup> For an arbitrary Hermitian operator  $\hat{O}$ , we can associate a standard deviation

$$\sigma_O = \sqrt{\langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2},$$

where the brackets  $\langle O \rangle$  indicate an expectation value. For a pair of operators  $\hat{A}$  and  $\hat{B}$ , we may define their *commutator* as

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A},$$

In this notation, the Robertson uncertainty relation is given by

$$\sigma_A \sigma_B \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right| = \frac{1}{2} \left| \langle [\hat{A}, \hat{B}] \rangle \right|.$$

The Robertson uncertainty relation immediately follows from a slightly stronger inequality, the *Schrödinger uncertainty relation*,<sup>[14]</sup>

$$\sigma_A^2 \sigma_B^2 \geq \left( \frac{1}{2} \langle \{\hat{A}, \hat{B}\} \rangle - \langle \hat{A} \rangle \langle \hat{B} \rangle \right)^2 + \left( \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right)^2,$$

where we have introduced the *anticommutator*,

$$\{\hat{A}, \hat{B}\} = \hat{A}\hat{B} + \hat{B}\hat{A}.$$

Proof of the Schrödinger uncertainty relation

The derivation shown here incorporates and builds off of those shown in Robertson,<sup>[13]</sup> Schrödinger<sup>[14]</sup> and standard textbooks such as Griffiths.<sup>[15]</sup> For any Hermitian operator  $\hat{A}$ , based upon the definition of variance, we have

$$\sigma_A^2 = \langle (\hat{A} - \langle \hat{A} \rangle) \Psi | (\hat{A} - \langle \hat{A} \rangle) \Psi \rangle.$$

we let  $|f\rangle = |(\hat{A} - \langle \hat{A} \rangle) \Psi\rangle$  and thus

$$\sigma_A^2 = \langle f | f \rangle.$$

Similarly, for any other Hermitian operator  $\hat{B}$  in the same state

$$\sigma_B^2 = \langle (\hat{B} - \langle \hat{B} \rangle) \Psi | (\hat{B} - \langle \hat{B} \rangle) \Psi \rangle = \langle g | g \rangle$$

for  $|g\rangle = |(\hat{B} - \langle \hat{B} \rangle) \Psi\rangle$ .

The product of the two deviations can thus be expressed as

{{{}}}

$$\sigma_A^2 \sigma_B^2 = \langle f | f \rangle \langle g | g \rangle. \quad (1)$$

{{{}}}

In order to relate the two vectors  $|f\rangle$  and  $|g\rangle$ , we use the Cauchy–Schwarz inequality<sup>[16]</sup> which is defined as

$$\langle f | f \rangle \langle g | g \rangle \geq |\langle f | g \rangle|^2,$$

and thus Eq. (1) can be written as

{{{}}}

$$\sigma_A^2 \sigma_B^2 \geq |\langle f | g \rangle|^2. \quad (2)$$

{{{}}}

Since  $\langle f | g \rangle$  is in general a complex number, we use the fact that the modulus squared of any complex number  $z$  is defined as  $|z|^2 = z z^*$ , where  $z^*$  is the complex conjugate of  $z$ . The modulus squared can also be expressed as

{{{}}}

$$|z|^2 = (\text{Re}(z))^2 + (\text{Im}(z))^2 = \left( \frac{z + z^*}{2} \right)^2 + \left( \frac{z - z^*}{2i} \right)^2. \quad (3)$$

{{{}}}

we let  $z = \langle f | g \rangle$  and  $z^* = \langle g | f \rangle$  and substitute these into the equation above to get

{{{}}}

$$|\langle f | g \rangle|^2 = \left( \frac{\langle f | g \rangle + \langle g | f \rangle}{2} \right)^2 + \left( \frac{\langle f | g \rangle - \langle g | f \rangle}{2i} \right)^2 \quad (4)$$

{{{}}}

The inner product  $\langle f|g\rangle$  is written out explicitly as

$$\langle f|g\rangle = \langle (\hat{A} - \langle \hat{A} \rangle)\Psi | (\hat{B} - \langle \hat{B} \rangle)\Psi \rangle,$$

and using the fact that  $\hat{A}$  and  $\hat{B}$  are Hermitian operators, we find

$$\begin{aligned} \langle f|g\rangle &= \langle \Psi | (\hat{A} - \langle \hat{A} \rangle)(\hat{B} - \langle \hat{B} \rangle)\Psi \rangle \\ &= \langle \Psi | (\hat{A}\hat{B} - \hat{A}\langle \hat{B} \rangle - \hat{B}\langle \hat{A} \rangle + \langle \hat{A} \rangle\langle \hat{B} \rangle)\Psi \rangle \\ &= \langle \Psi | \hat{A}\hat{B}\Psi \rangle - \langle \Psi | \hat{A}\langle \hat{B} \rangle\Psi \rangle - \langle \Psi | \hat{B}\langle \hat{A} \rangle\Psi \rangle + \langle \Psi | \langle \hat{A} \rangle\langle \hat{B} \rangle\Psi \rangle \\ &= \langle \hat{A}\hat{B} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle + \langle \hat{A} \rangle\langle \hat{B} \rangle \\ &= \langle \hat{A}\hat{B} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle. \end{aligned}$$

Similarly it can be shown that  $\langle g|f\rangle = \langle \hat{B}\hat{A} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle$ . Thus we have

$$\langle f|g\rangle - \langle g|f\rangle = \langle \hat{A}\hat{B} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle - \langle \hat{B}\hat{A} \rangle + \langle \hat{A} \rangle\langle \hat{B} \rangle = \langle [\hat{A}, \hat{B}] \rangle$$

and

$$\langle f|g\rangle + \langle g|f\rangle = \langle \hat{A}\hat{B} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle + \langle \hat{B}\hat{A} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle = \langle \{\hat{A}, \hat{B}\} \rangle - 2\langle \hat{A} \rangle\langle \hat{B} \rangle.$$

We now substitute the above two equations above back into Eq. (4) and get

$$|\langle f|g\rangle|^2 = \left(\frac{1}{2}\langle \{\hat{A}, \hat{B}\} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle\right)^2 + \left(\frac{1}{2i}\langle [\hat{A}, \hat{B}] \rangle\right)^2.$$

Substituting the above into Eq. (2) we get the Schrödinger uncertainty relation

$$\sigma_A\sigma_B \geq \sqrt{\left(\frac{1}{2}\langle \{\hat{A}, \hat{B}\} \rangle - \langle \hat{A} \rangle\langle \hat{B} \rangle\right)^2 + \left(\frac{1}{2i}\langle [\hat{A}, \hat{B}] \rangle\right)^2}.$$

This proof has a small issue related to the Hermiticity of  $A$  when the state is an eigenstate of  $B$ .<sup>[17]</sup> This issue can be overcome by using a variational method for the proof.<sup>[18][19]</sup>

Since the Robertson and Schrödinger relations are for general operators, the relations can be applied to any two observables to obtain specific uncertainty relations. A few of the most common relations found in the literature are given below.

- For position and linear momentum, the canonical commutation relation  $[\hat{x}, \hat{p}] = i\hbar$  implies the Kennard inequality from above:

$$\sigma_x\sigma_p \geq \frac{\hbar}{2}$$

- For two orthogonal components of the total angular momentum operator of an object:

$$\sigma_{J_i}\sigma_{J_j} \geq \frac{\hbar}{2} |\langle J_k \rangle|,$$

where  $i, j, k$  are distinct and  $J_i$  denotes angular momentum along the  $x_i$  axis. This relation implies that unless all three components vanish together, only a single component of a system's angular momentum can be defined with arbitrary precision, normally the component parallel to an external (magnetic or electric) field. Moreover, for  $[J_x, J_y] = i\hbar\epsilon_{xyz}J_z$ , a choice  $\hat{A} = J_x, \hat{B} = J_y$ , in angular momentum multiplets,  $\psi = |j, m\rangle$ , bounds the Casimir invariant (angular momentum squared,  $\langle J_x^2 + J_y^2 + J_z^2 \rangle$ ) from below and thus yields useful constraints such as  $j(j+1) \geq m(m+1)$ , and hence  $j \geq m$ , among others.

- In non-relativistic mechanics, time is privileged as an independent variable. Nevertheless, in 1945, L. I. Mandelshtam and I. E. Tamm derived a non-relativistic time-energy uncertainty relation, as follows.<sup>[20]</sup> For a quantum system in a non-stationary state  $\psi$  and an observable  $B$  represented by a self-adjoint operator  $\hat{B}$ , the following formula holds:



$$\sigma_E \frac{\sigma_B}{\left| \frac{d\langle \hat{B} \rangle}{dt} \right|} \geq \frac{\hbar}{2},$$

where  $\sigma_E$  is the standard deviation of the energy operator in the state  $\psi$ ,  $\sigma_B$  stands for the standard deviation of  $B$ . Although the second factor in the left-hand side has dimension of time, it is different from the time parameter that enters Schrödinger equation. It is a *lifetime* of the state  $\psi$  with respect to the observable  $B$ . In other words, this is the time after which the expectation value  $\langle \hat{B} \rangle$  changes appreciably.

- For the number of electrons in a superconductor and the phase of its Ginzburg–Landau order parameter<sup>[21][22]</sup>

$$\Delta N \Delta \phi \geq 1 .$$

## Examples

### Quantum harmonic oscillator stationary states

Consider a one-dimensional quantum harmonic oscillator (QHO). It is possible to express the position and momentum operators in terms of the creation and annihilation operators:

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega}}(a + a^\dagger)$$

$$\hat{p} = i\sqrt{\frac{m\omega\hbar}{2}}(a^\dagger - a).$$

Using the rules for creation and annihilation operators on the eigenstates of the QHO,

$$a^\dagger |n\rangle = \sqrt{n+1} |n+1\rangle$$

$$a |n\rangle = \sqrt{n} |n-1\rangle,$$

we may show that the variances are

$$\sigma_x^2 = \frac{\hbar}{m\omega} \left( n + \frac{1}{2} \right)$$

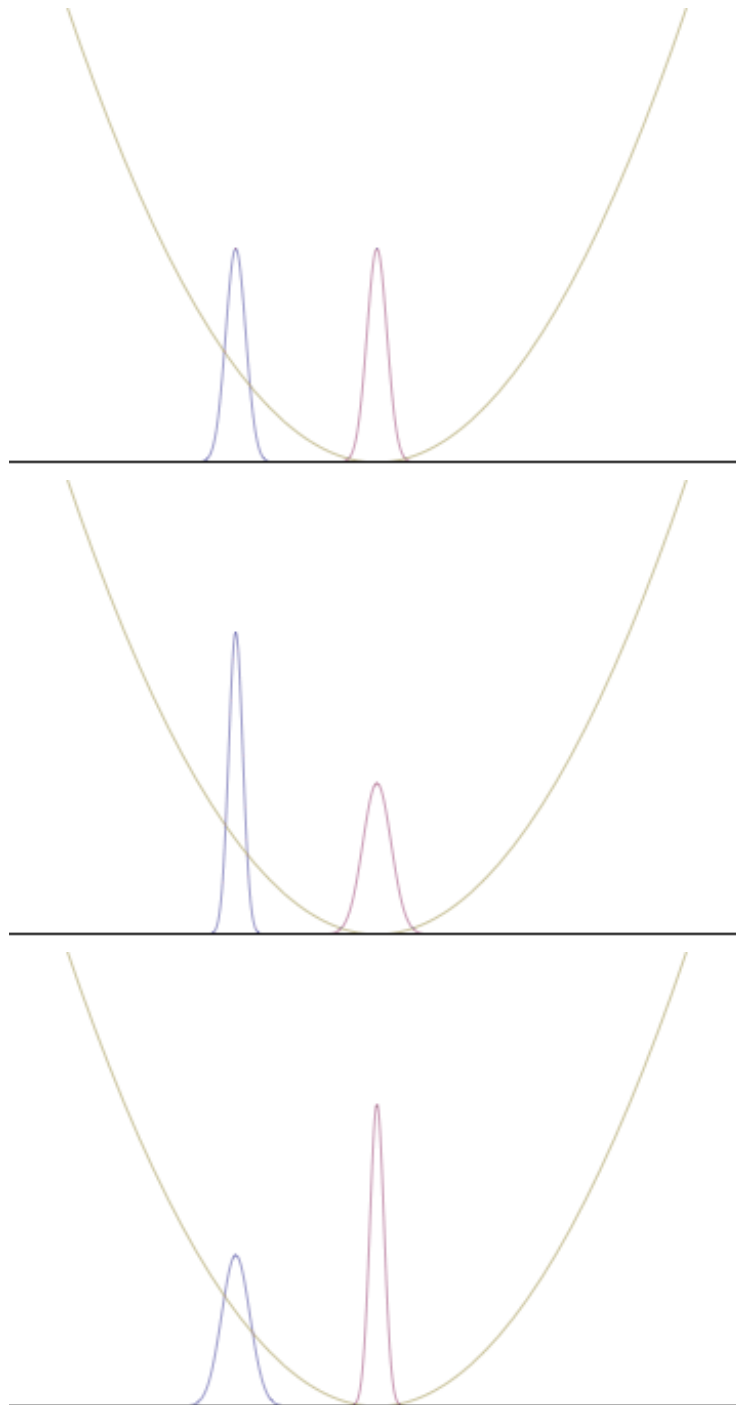
$$\sigma_p^2 = \hbar m\omega \left( n + \frac{1}{2} \right).$$

The product of these standard deviations is therefore

$$\sigma_x \sigma_p = \hbar \left( n + \frac{1}{2} \right) \geq \frac{\hbar}{2}.$$

In particular, the Kennard bound<sup>[2]</sup> is saturated for the ground state  $n = 0$ , where the probability density is the normal distribution.

### Quantum harmonic oscillator with Gaussian initial condition



In a quantum harmonic oscillator of characteristic angular frequency  $\omega$ , place a state that is offset from the bottom of the potential by some displacement  $x_0$  as

$$\psi(x) = \left(\frac{m\Omega}{\pi\hbar}\right)^{1/4} \exp\left(-\frac{m\Omega(x-x_0)^2}{2\hbar}\right),$$

where  $\Omega$  describes the width of the initial state but need not be the same as  $\omega$ . Through integration over the propagator, we can solve for the full time-dependent solution. After many cancelations, the probability densities reduce to

$$|\Psi(x, t)|^2 \sim \mathcal{N}\left(x_0 \cos(\omega t), \frac{\hbar}{2m\Omega} \left(\cos^2(\omega t) + \frac{\Omega^2}{\omega^2} \sin^2(\omega t)\right)\right)$$

$$|\Phi(p, t)|^2 \sim \mathcal{N}\left(-mx_0\omega \sin(\omega t), \frac{\hbar m\Omega}{2} \left(\cos^2(\omega t) + \frac{\omega^2}{\Omega^2} \sin^2(\omega t)\right)\right),$$

where we have used the notation  $\mathcal{N}(\mu, \sigma^2)$  to denote a normal distribution of mean  $\mu$  and variance  $\sigma^2$ . Copying the variances above and applying trigonometric identities, we can write the product of the standard deviations as

$$\begin{aligned} \sigma_x \sigma_p &= \frac{\hbar}{2} \sqrt{\left(\cos^2(\omega t) + \frac{\Omega^2}{\omega^2} \sin^2(\omega t)\right) \left(\cos^2(\omega t) + \frac{\omega^2}{\Omega^2} \sin^2(\omega t)\right)} \\ &= \frac{\hbar}{4} \sqrt{3 + \frac{1}{2} \left(\frac{\Omega^2}{\omega^2} + \frac{\omega^2}{\Omega^2}\right) - \left(\frac{1}{2} \left(\frac{\Omega^2}{\omega^2} + \frac{\omega^2}{\Omega^2}\right) - 1\right) \cos(4\omega t)} \end{aligned}$$

From the relations

$$\frac{\Omega^2}{\omega^2} + \frac{\omega^2}{\Omega^2} \geq 2, \quad |\cos(4\omega t)| \leq 1,$$

we can conclude

$$\sigma_x \sigma_p \geq \frac{\hbar}{4} \sqrt{3 + \frac{1}{2} \left(\frac{\Omega^2}{\omega^2} + \frac{\omega^2}{\Omega^2}\right) - \left(\frac{1}{2} \left(\frac{\Omega^2}{\omega^2} + \frac{\omega^2}{\Omega^2}\right) - 1\right)} = \frac{\hbar}{2}.$$

### Coherent states

A coherent state is a right eigenstate of the annihilation operator,

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle,$$

which may be represented in terms of Fock states as

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$$

In the picture where the coherent state is a massive particle in a QHO, the position and momentum operators may be expressed in terms of the annihilation operators in the same formulas above and used to calculate the variances,

$$\begin{aligned} \sigma_x^2 &= \frac{\hbar}{2m\omega} \\ \sigma_p^2 &= \frac{\hbar m\omega}{2}. \end{aligned}$$

Therefore every coherent state saturates the Kennard bound

$$\sigma_x \sigma_p = \sqrt{\frac{\hbar}{2m\omega}} \sqrt{\frac{\hbar m\omega}{2}} = \frac{\hbar}{2}.$$

with position and momentum each contributing an amount  $\sqrt{\hbar/2}$  in a "balanced" way. Moreover every squeezed coherent state also saturates the Kennard bound although the individual contributions of position and momentum need not be balanced in general.

### Particle in a box

Consider a particle in a one-dimensional box of length  $L$ . The eigenfunctions in position and momentum space are

$$\psi_n(x, t) = \begin{cases} A \sin(k_n x) e^{-i\omega_n t}, & 0 < x < L, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\phi_n(p, t) = \sqrt{\frac{\pi L}{\hbar}} \frac{n \left(1 - (-1)^n e^{-ikL}\right) e^{-i\omega_n t}}{\pi^2 n^2 - k^2 L^2},$$

where  $\omega_n = \frac{\pi^2 \hbar n^2}{8L^2 m}$  and we have used the de Broglie relation  $p = \hbar k$ . The variances of  $x$  and  $p$  can be calculated explicitly:

$$\sigma_x^2 = \frac{L^2}{12} \left(1 - \frac{6}{n^2 \pi^2}\right)$$

$$\sigma_p^2 = \left(\frac{\hbar n \pi}{L}\right)^2.$$

The product of the standard deviations is therefore

$$\sigma_x \sigma_p = \frac{\hbar}{2} \sqrt{\frac{n^2 \pi^2}{3} - 2}.$$

For all  $n = 1, 2, 3, \dots$ , the quantity  $\sqrt{\frac{n^2 \pi^2}{3} - 2}$  is greater than 1, so the uncertainty principle is never violated.

For numerical concreteness, the smallest value occurs when  $n = 1$ , in which case

$$\sigma_x \sigma_p = \frac{\hbar}{2} \sqrt{\frac{\pi^2}{3} - 2} \approx 0.568 \hbar > \frac{\hbar}{2}.$$

### Constant momentum

Assume a particle initially has a momentum space wave function described by a normal distribution around some constant momentum  $p_0$  according to

$$\phi(p) = \left(\frac{x_0}{\hbar \sqrt{\pi}}\right)^{1/2} \cdot \exp\left(\frac{-x_0^2 (p - p_0)^2}{2\hbar^2}\right),$$

where we have introduced a reference scale  $x_0 = \sqrt{\hbar/m\omega_0}$ , with  $\omega_0 > 0$  describing the width of the distribution—cf. nondimensionalization. If the state is allowed to evolve in free space, then the time-dependent momentum and position space wave functions are

$$\Phi(p, t) = \left(\frac{x_0}{\hbar \sqrt{\pi}}\right)^{1/2} \cdot \exp\left(\frac{-x_0^2 (p - p_0)^2}{2\hbar^2} - \frac{ip^2 t}{\hbar}\right),$$

$$\Psi(x, t) = \left(\frac{1}{x_0 \sqrt{\pi}}\right)^{1/2} \cdot \frac{e^{-x_0^2 p_0^2 / 2\hbar^2}}{\sqrt{1 + i\omega_0 t}} \cdot \exp\left(-\frac{(x - ix_0^2 p_0 / \hbar)^2}{2x_0^2 (1 + i\omega_0 t)}\right).$$

Since  $\langle p(t) \rangle = p_0$  and  $\sigma_p(t) = \hbar/x_0 \sqrt{2}$ , this can be interpreted as a particle moving along with constant momentum at arbitrarily high precision. On the other hand, the standard deviation of the position is

$$\sigma_x = \frac{x_0}{\sqrt{2}} \sqrt{1 + \omega_0^2 t^2}$$

such that the uncertainty product can only increase with time as

$$\sigma_x(t)\sigma_p(t) = \frac{\hbar}{2}\sqrt{1 + \omega_0^2 t^2}$$

### Additional uncertainty relations

#### Mixed states

The Robertson–Schrödinger uncertainty relation may be generalized in a straightforward way to describe mixed states.<sup>[23]</sup>

$$\sigma_A^2\sigma_B^2 \geq \left(\frac{1}{2}\text{tr}(\rho\{A, B\}) - \text{tr}(\rho A)\text{tr}(\rho B)\right)^2 + \left(\frac{1}{2i}\text{tr}(\rho[A, B])\right)^2$$

#### Phase space

In the phase space formulation of quantum mechanics, the Robertson–Schrödinger relation follows from a positivity condition on a real star-square function. Given a Wigner function  $W(x, p)$  with star product  $\star$  and a function  $f$ , the following is generally true<sup>[24]</sup>:

$$\langle f^* \star f \rangle = \int (f^* \star f) W(x, p) dx dp \geq 0.$$

Choosing  $f = a + bx + cp$ , we arrive at

$$\langle f^* \star f \rangle = \begin{bmatrix} a^* & b^* & c^* \end{bmatrix} \begin{bmatrix} 1 & \langle x \rangle & \langle p \rangle \\ \langle x \rangle & \langle x \star x \rangle & \langle x \star p \rangle \\ \langle p \rangle & \langle p \star x \rangle & \langle p \star p \rangle \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \geq 0.$$

Since this positivity condition is true for *all*  $a, b$ , and  $c$ , it follows that all the eigenvalues of the matrix are positive. The positive eigenvalues then imply a corresponding positivity condition on the determinant:

$$\det \begin{bmatrix} 1 & \langle x \rangle & \langle p \rangle \\ \langle x \rangle & \langle x \star x \rangle & \langle x \star p \rangle \\ \langle p \rangle & \langle p \star x \rangle & \langle p \star p \rangle \end{bmatrix} = \det \begin{bmatrix} 1 & \langle x \rangle & \langle p \rangle \\ \langle x \rangle & \langle x^2 \rangle & \langle xp + \frac{i\hbar}{2} \rangle \\ \langle p \rangle & \langle xp - \frac{i\hbar}{2} \rangle & \langle p^2 \rangle \end{bmatrix} \geq 0,$$

or, explicitly, after algebraic manipulation,

$$\sigma_x^2\sigma_p^2 = (\langle x^2 \rangle - \langle x \rangle^2) (\langle p^2 \rangle - \langle p \rangle^2) \geq (\langle xp \rangle - \langle x \rangle \langle p \rangle)^2 + \frac{\hbar^2}{4}.$$

#### Systematic error

The inequalities above focus on the statistical imprecision of observables as quantified by the standard deviation. Heisenberg's original version, however, was interested in *systematic error*, incurred by a disturbance of a quantum system by the measuring apparatus, i.e., an observer effect. If we let  $\epsilon_{\mathcal{O}}$  represent the error (i.e., accuracy) of a measurement of an observable  $\mathcal{O}$  and  $\eta_{\mathcal{O}}$  represent its disturbance by the measurement process, then the following inequality holds:<sup>[5]</sup>

$$\epsilon_A\eta_B + \epsilon_A\sigma_B + \sigma_A\eta_B \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right|$$

In fact, Heisenberg's uncertainty principle as originally described in the 1927 formulation mentions only the first term. Applying the notation above to Heisenberg's position-momentum relation, Heisenberg's argument could be rewritten as

$$\cancel{\epsilon_x\eta_p} \sim \frac{\hbar}{2} \text{ (Heisenberg).}$$

Such a formulation is both mathematically incorrect and experimentally refuted.<sup>[25]</sup> It is also possible to derive a similar uncertainty relation combining both the statistical and systematic error components.<sup>[26]</sup>

### Entropic uncertainty principle

For many distributions, the standard deviation is not a particularly natural way of quantifying the structure. For example, uncertainty relations in which one of the observables is an angle has little physical meaning for fluctuations larger than one period.<sup>[19][27][28][29]</sup> Other examples include highly bimodal distributions, or unimodal distributions with divergent variance.

A solution that overcomes these issues is an uncertainty based on entropic uncertainty instead of the product of variances. While formulating the many-worlds interpretation of quantum mechanics in 1957, Hugh Everett III conjectured a stronger extension of the uncertainty principle based on entropic certainty.<sup>[30]</sup> This conjecture, also studied by Hirschman<sup>[31]</sup> and proven in 1975 by Beckner<sup>[32]</sup> and by Iwo Białynicki-Birula and Jerzy Mycielski<sup>[33]</sup> is

$$H_x + H_p \geq \ln(e\pi)$$

where we have used the Shannon entropy (*not* the quantum von Neumann entropy)

$$H_x = - \int |\psi(x)|^2 \ln(|\psi(x)|^2 \cdot \ell) dx = - \langle \ln(|\psi(x)|^2 \cdot \ell) \rangle$$

$$H_p = - \int |\phi(p)|^2 \ln(|\phi(p)|^2 \cdot \hbar/\ell) dp = - \langle \ln(|\phi(p)|^2 \cdot \hbar/\ell) \rangle$$

for some arbitrary fixed length scale  $\ell$ .

From the inverse logarithmic Sobolev inequalities<sup>[34]</sup>

$$H_x \leq \frac{1}{2} \ln(2e\pi\sigma_x^2/\ell^2) ,$$

$$H_p \leq \frac{1}{2} \ln(2e\pi\sigma_p^2\ell^2/\hbar^2) ,$$

(equivalently, from the fact that normal distributions maximize the entropy of all such with a given variance), it readily follows that this entropic uncertainty principle is *stronger than the one based on standard deviations*, because

$$\sigma_x\sigma_p \geq \frac{\hbar}{2} \cdot \exp(H_x + H_p - \ln(e\pi)) \geq \frac{\hbar}{2} .$$

A few remarks on these inequalities. First, the choice of base e is a matter of popular convention in physics. The logarithm can alternatively be in any base, provided that it be consistent on both sides of the inequality. Second, the numerical value on the right hand side assumes the unitary convention of the Fourier transform, used throughout physics and elsewhere in this article. Third, the normal distribution saturates the inequality, and it is the only distribution with this property, because it is the maximum entropy probability distribution among those with fixed variance (cf. here for proof).

**Entropic uncertainty of the normal distribution**

We demonstrate this method on the ground state of the QHO, which as discussed above saturates the usual uncertainty based on standard deviations. The length scale can be set to whatever is convenient, so we assign

$$\begin{aligned}\ell &= \sqrt{\frac{\hbar}{2m\omega}} \\ \psi(x) &= \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \exp\left(-\frac{m\omega x^2}{2\hbar}\right) \\ &= \left(\frac{1}{2\pi\ell^2}\right)^{1/4} \exp\left(-\frac{x^2}{4\ell^2}\right)\end{aligned}$$

The probability distribution is the normal distribution

$$|\psi(x)|^2 = \frac{1}{\ell\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\ell^2}\right)$$

with Shannon entropy

$$\begin{aligned}H_x &= - \int |\psi(x)|^2 \ln(|\psi(x)|^2 \cdot \ell) dx \\ &= -\frac{1}{\ell\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\ell^2}\right) \ln\left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\ell^2}\right)\right] dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) \left[\ln(\sqrt{2\pi}) + \frac{u^2}{2}\right] du \\ &= \ln(\sqrt{2\pi}) + \frac{1}{2}.\end{aligned}$$

A completely analogous calculation proceeds for the momentum distribution.

$$\begin{aligned}\phi(p) &= \left(\frac{2\ell^2}{\pi\hbar^2}\right)^{1/4} \exp\left(-\frac{\ell^2 p^2}{\hbar^2}\right) \\ |\phi(p)|^2 &= \sqrt{\frac{2\ell^2}{\pi\hbar^2}} \exp\left(-\frac{2\ell^2 p^2}{\hbar^2}\right) \\ H_p &= - \int |\phi(p)|^2 \ln(|\phi(p)|^2 \cdot \hbar/\ell) dp \\ &= -\sqrt{\frac{2\ell^2}{\pi\hbar^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{2\ell^2 p^2}{\hbar^2}\right) \ln\left[\sqrt{\frac{2}{\pi}} \exp\left(-\frac{2\ell^2 p^2}{\hbar^2}\right)\right] dp \\ &= \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} \exp(-2v^2) \left[\ln\left(\sqrt{\frac{\pi}{2}}\right) + 2v^2\right] dv \\ &= \ln\left(\sqrt{\frac{\pi}{2}}\right) + \frac{1}{2}.\end{aligned}$$

The entropic uncertainty is therefore the limiting value

$$\begin{aligned}H_x + H_p &= \ln(\sqrt{2\pi}) + \frac{1}{2} + \ln\left(\sqrt{\frac{\pi}{2}}\right) + \frac{1}{2} \\ &= 1 + \ln \pi = \ln(e\pi).\end{aligned}$$



### Harmonic analysis

In the context of harmonic analysis, a branch of mathematics, the uncertainty principle implies that one cannot at the same time localize the value of a function and its Fourier transform. To wit, the following inequality holds:

$$\left( \int_{-\infty}^{\infty} x^2 |f(x)|^2 dx \right) \left( \int_{-\infty}^{\infty} \xi^2 |\hat{f}(\xi)|^2 d\xi \right) \geq \frac{\|f\|_2^4}{16\pi^2}.$$

Other purely mathematical formulations of uncertainty exist between a function  $f$  and its Fourier transform.<sup>[35][36][37]</sup>

### Signal processing

In the context of signal processing, particularly time–frequency analysis, uncertainty principles are referred to as the **Gabor limit**, after Dennis Gabor, or sometimes the *Heisenberg–Gabor limit*. The basic result, which follows from Benedicks's theorem, below, is that a function cannot be both time limited and band limited (a function and its Fourier transform cannot both have bounded domain) – see bandlimited versus timelimited. Stated alternatively, "one cannot simultaneously localize a signal (function) in both the time domain ( $f$ ) and frequency domain (Fourier transform)". When applied to filters, the result is that one cannot achieve high temporal resolution and frequency resolution at the same time; a concrete example are the resolution issues of the short-time Fourier transform – if one uses a wide window, one achieves good frequency resolution at the cost of temporal resolution, while a narrow window has the opposite trade-off.

Alternative theorems give more precise quantitative results, and in time–frequency analysis, rather than interpreting the (1-dimensional) time and frequency domains separately, one instead interprets the limit as a lower limit on the support of a function in the (2-dimensional) time–frequency plane. In practice the Gabor limit limits the *simultaneous* time–frequency resolution one can achieve without interference; it is possible to achieve higher resolution, but at the cost of different components of the signal interfering with each other.

### Benedicks's theorem

Amrein-Berthier<sup>[38]</sup> and Benedicks's theorem<sup>[39]</sup> intuitively says that the set of points where  $f$  is non-zero and the set of points where  $\hat{f}$  is nonzero cannot both be small. Specifically, it is impossible for a function  $f$  in  $L^2(\mathbf{R})$  and its Fourier transform to both be supported on sets of finite Lebesgue measure. A more quantitative version is due to Nazarov.<sup>[40][41]</sup>

$$\|f\|_{L^2(\mathbf{R}^d)} \leq C e^{C|S||\Sigma|} (\|f\|_{L^2(S^c)} + \|\hat{f}\|_{L^2(\Sigma^c)})$$

One expects that the factor  $C e^{C|S||\Sigma|}$  may be replaced by  $C e^{C(|S||\Sigma|)^{1/d}}$  which is only known if either  $S$  or  $\Sigma$  is convex.

### Hardy's uncertainty principle

The mathematician G. H. Hardy formulated the following uncertainty principle:<sup>[42]</sup> it is not possible for  $f$  and  $\hat{f}$  to both be "very rapidly decreasing." Specifically, if  $f$  is in  $L^2(\mathbf{R})$ , is such that

$$|f(x)| \leq C(1 + |x|)^N e^{-a\pi x^2}$$

and

$$|\hat{f}(\xi)| \leq C(1 + |\xi|)^N e^{-b\pi \xi^2} \quad (C > 0, N \text{ an integer})$$

then, if  $ab > 1$ ,  $f = 0$  while if  $ab = 1$  then there is a polynomial  $P$  of degree  $\leq N$  such that

$$f(x) = P(x)e^{-a\pi x^2}.$$

This was later improved as follows: if  $f \in L^2(\mathbf{R}^d)$  is such that

$$\int_{\mathbf{R}^d} \int_{\mathbf{R}^d} |f(x)| |\hat{f}(\xi)| \frac{e^{\pi|(x,\xi)|}}{(1 + |x| + |\xi|)^N} dx d\xi < +\infty$$

then

$$f(x) = P(x)e^{-\pi(Ax,x)}$$

where  $P$  is a polynomial of degree  $< \frac{N-d}{2}$  and  $A$  is a real  $d \times d$  positive definite matrix.

This result was stated in Beurling's complete works without proof and proved in Hörmander<sup>[43]</sup> (the case  $d = 1, N = 0$ ) and Bonami, Demange, and Jaming<sup>[44]</sup> for the general case. Note that Hörmander–Beurling's version implies the case  $ab > 1$  in Hardy's Theorem while the version by Bonami–Demange–Jaming covers the full strength of Hardy's Theorem.

A full description of the case  $ab < 1$  as well as the following extension to Schwarz class distributions appears in Demange:<sup>[45]</sup>

**Theorem.** If a tempered distribution  $f \in \mathcal{S}'(\mathbb{R}^d)$  is such that

$$e^{\pi|x|^2} f \in \mathcal{S}'(\mathbb{R}^d)$$

and

$$e^{\pi|\xi|^2} \hat{f} \in \mathcal{S}'(\mathbb{R}^d)$$

then

$$f(x) = P(x)e^{-\pi(Ax,x)}$$

for some convenient polynomial  $P$  and real positive definite matrix  $A$  of type  $d \times d$ .

## History

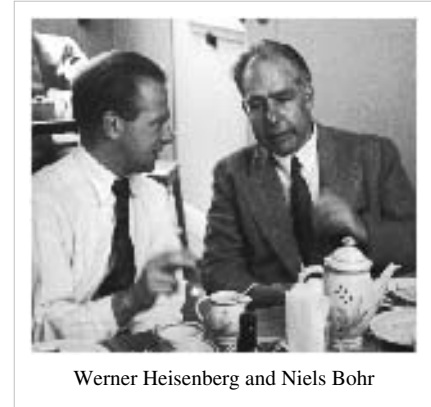
Werner Heisenberg formulated the Uncertainty Principle at Niels Bohr's institute in Copenhagen, while working on the mathematical foundations of quantum mechanics.<sup>[46]</sup>

In 1925, following pioneering work with Hendrik Kramers, Heisenberg developed matrix mechanics, which replaced the ad-hoc old quantum theory with modern quantum mechanics. The central assumption was that the classical concept of motion does not fit at the quantum level, and that electrons in an atom do not travel on sharply defined orbits. Rather, the motion is smeared out in a strange way: the Fourier transform of time only involve those frequencies that could be seen in quantum jumps.

Heisenberg's paper did not admit any unobservable quantities like the exact position of the electron in an orbit at any time; he only allowed the theorist to talk about the Fourier components of the motion. Since the Fourier components were not defined at the classical frequencies, they could not be used to construct an exact trajectory, so that the formalism could not answer certain overly precise questions about where the electron was or how fast it was going.

In March 1926, working in Bohr's institute, Heisenberg realized that the non-commutativity implies the uncertainty principle. This implication provided a clear physical interpretation for the non-commutativity, and it laid the foundation for what became known as the Copenhagen interpretation of quantum mechanics. Heisenberg showed that the commutation relation implies an uncertainty, or in Bohr's language a complementarity.<sup>[47]</sup> Any two variables that do not commute cannot be measured simultaneously—the more precisely one is known, the less precisely the other can be known. Heisenberg wrote:

It can be expressed in its simplest form as follows: One can never know with perfect accuracy both of those two important factors which determine the movement of one of the smallest particles—its position and its velocity. It is impossible to determine accurately *both* the position and the direction and speed of



Werner Heisenberg and Niels Bohr

a particle *at the same instant*.<sup>[48]</sup>

In his celebrated 1927 paper, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik" ("On the Perceptual Content of Quantum Theoretical Kinematics and Mechanics"), Heisenberg established this expression as the minimum amount of unavoidable momentum disturbance caused by any position measurement,<sup>[49]</sup> but he did not give a precise definition for the uncertainties  $\Delta x$  and  $\Delta p$ . Instead, he gave some plausible estimates in each case separately. In his Chicago lecture<sup>[50]</sup> he refined his principle:

$$\Delta x \Delta p \gtrsim h \quad (1)$$

Kennard<sup>[2]</sup> in 1927 first proved the modern inequality:

$$\sigma_x \sigma_p \geq \frac{\hbar}{2} \quad (2)$$

where  $\hbar = h/2\pi$ , and  $\sigma_x, \sigma_p$  are the standard deviations of position and momentum. Heisenberg only proved relation (2) for the special case of Gaussian states.<sup>[50]</sup>

## Terminology and translation

Throughout the main body of his original 1927 paper, written in German, Heisenberg used the word, "Unbestimmtheit" ("indeterminacy"), to describe the basic theoretical principle. Only in the endnote did he switch to the word, "Unsicherheit" ("uncertainty"). When the English-language version of Heisenberg's textbook, *The Physical Principles of the Quantum Theory*, was published in 1930, however, the translation "uncertainty" was used, and it became the more commonly used term in the English language thereafter.<sup>[51]</sup>

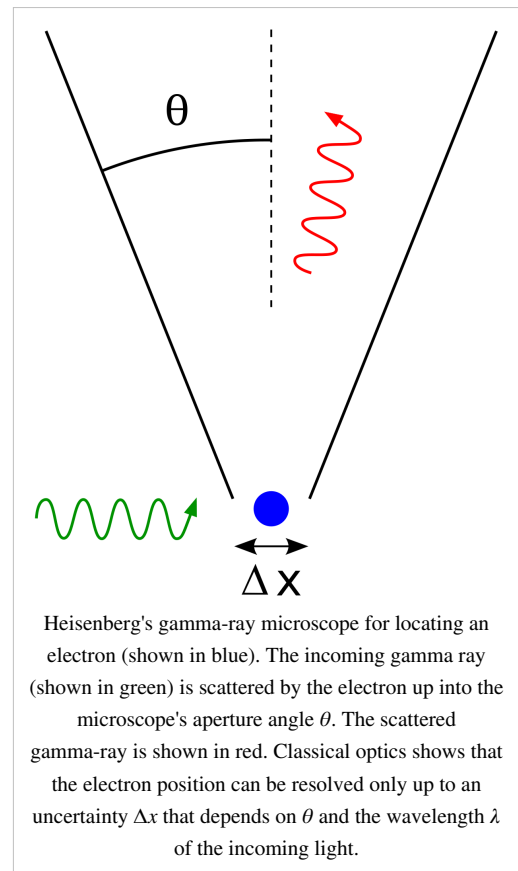
## Heisenberg's microscope

The principle is quite counter-intuitive, so the early students of quantum theory had to be reassured that naive measurements to violate it, were bound always to be unworkable. One way in which Heisenberg originally illustrated the intrinsic impossibility of violating the uncertainty principle is by using an imaginary microscope as a measuring device.<sup>[50]</sup>

He imagines an experimenter trying to measure the position and momentum of an electron by shooting a photon at it.

**Problem 1** – If the photon has a short wavelength, and therefore, a large momentum, the position can be measured accurately. But the photon scatters in a random direction, transferring a large and uncertain amount of momentum to the electron. If the photon has a long wavelength and low momentum, the collision does not disturb the electron's momentum very much, but the scattering will reveal its position only vaguely.

**Problem 2** – If a large aperture is used for the microscope, the electron's location can be well resolved (see Rayleigh criterion); but by the principle of conservation of momentum, the transverse momentum of the incoming photon and hence, the new momentum of the electron resolves poorly. If a small aperture is used, the accuracy of both resolutions is the other way around.



The combination of these trade-offs imply that no matter what photon wavelength and aperture size are used, the product of the uncertainty in measured position and measured momentum is greater than or equal to a lower limit, which is (up to a small numerical factor) equal to Planck's constant.<sup>[52]</sup> Heisenberg did not care to formulate the uncertainty principle as an exact limit (which is elaborated below), and preferred to use it instead, as a heuristic quantitative statement, correct up to small numerical factors, which makes the radically new noncommutativity of quantum mechanics inevitable.

## Critical reactions

The Copenhagen interpretation of quantum mechanics and Heisenberg's Uncertainty Principle were, in fact, seen as twin targets by detractors who believed in an underlying determinism and realism. According to the Copenhagen interpretation of quantum mechanics, there is no fundamental reality that the quantum state describes, just a prescription for calculating experimental results. There is no way to say what the state of a system fundamentally is, only what the result of observations might be.

Albert Einstein believed that randomness is a reflection of our ignorance of some fundamental property of reality, while Niels Bohr believed that the probability distributions are fundamental and irreducible, and depend on which measurements we choose to perform. Einstein and Bohr debated the uncertainty principle for many years. Some experiments within the first decade of the twenty-first century have cast doubt on how extensively the uncertainty principle applies.<sup>[53]</sup>

## Einstein's slit

The first of Einstein's thought experiments challenging the uncertainty principle went as follows:

Consider a particle passing through a slit of width  $d$ . The slit introduces an uncertainty in momentum of approximately  $h/d$  because the particle passes through the wall. But let us determine the momentum of the particle by measuring the recoil of the wall. In doing so, we find the momentum of the particle to arbitrary accuracy by conservation of momentum.

Bohr's response was that the wall is quantum mechanical as well, and that to measure the recoil to accuracy  $\Delta p$  the momentum of the wall must be known to this accuracy before the particle passes through. This introduces an uncertainty in the position of the wall and therefore the position of the slit equal to  $h/\Delta p$ , and if the wall's momentum is known precisely enough to measure the recoil, the slit's position is uncertain enough to disallow a position measurement.

A similar analysis with particles diffracting through multiple slits is given by Richard Feynman.<sup>[54]</sup>

## Einstein's box

Bohr was present when Einstein proposed the thought experiment which has become known as Einstein's box. Einstein argued that "Heisenberg's uncertainty equation implied that the uncertainty in time was related to the uncertainty in energy, the product of the two being related to Planck's constant."<sup>[55]</sup> Consider, he said, an ideal box, lined with mirrors so that it can contain light indefinitely. The box could be weighed before a clockwork mechanism opened an ideal shutter at a chosen instant to allow one single photon to escape. "We now know, explained Einstein, precisely the time at which the photon left the box."<sup>[56]</sup> "Now, weigh the box again. The change of mass tells the energy of the emitted light. In this manner, said Einstein, one could measure the energy emitted and the time it was released with any desired precision, in contradiction to the uncertainty principle."<sup>[55]</sup>

Bohr spent a sleepless night considering this argument, and eventually realized that it was flawed. He pointed out that if the box were to be weighed, say by a spring and a pointer on a scale, "since the box must move vertically with a change in its weight, there will be uncertainty in its vertical velocity and therefore an uncertainty in its height above the table. ... Furthermore, the uncertainty about the elevation above the earth's surface will result in an

uncertainty in the rate of the clock,<sup>[57]</sup> because of Einstein's own theory of gravity's effect on time. "Through this chain of uncertainties, Bohr showed that Einstein's light box experiment could not simultaneously measure exactly both the energy of the photon and the time of its escape."<sup>[58]</sup>

### EPR paradox for entangled particles

Bohr was compelled to modify his understanding of the uncertainty principle after another thought experiment by Einstein. In 1935, Einstein, Podolsky and Rosen (see EPR paradox) published an analysis of widely separated entangled particles. Measuring one particle, Einstein realized, would alter the probability distribution of the other, yet here the other particle could not possibly be disturbed. This example led Bohr to revise his understanding of the principle, concluding that the uncertainty was not caused by a direct interaction.<sup>[59]</sup>

But Einstein came to much more far-reaching conclusions from the same thought experiment. He believed the "natural basic assumption" that a complete description of reality, would have to predict the results of experiments from "locally changing deterministic quantities", and therefore, would have to include more information than the maximum possible allowed by the uncertainty principle.

In 1964, John Bell showed that this assumption can be falsified, since it would imply a certain inequality between the probabilities of different experiments. Experimental results confirm the predictions of quantum mechanics, ruling out Einstein's basic assumption that led him to the suggestion of his *hidden variables*. Ironically this fact is one of the best pieces of evidence supporting Karl Popper's philosophy of invalidation of a theory by falsification-experiments. That is to say, here Einstein's "basic assumption" became falsified by experiments based on Bell's inequalities. For the objections of Karl Popper against the Heisenberg inequality itself, see below.

While it is possible to assume that quantum mechanical predictions are due to nonlocal, hidden variables, and in fact David Bohm invented such a formulation, this resolution is not satisfactory to the vast majority of physicists. The question of whether a random outcome is predetermined by a nonlocal theory can be philosophical, and it can be potentially intractable. If the hidden variables are not constrained, they could just be a list of random digits that are used to produce the measurement outcomes. To make it sensible, the assumption of nonlocal hidden variables is sometimes augmented by a second assumption—that the size of the observable universe puts a limit on the computations that these variables can do. A nonlocal theory of this sort predicts that a quantum computer would encounter fundamental obstacles when attempting to factor numbers of approximately 10,000 digits or more; a potentially achievable task in quantum mechanics.<sup>[60]</sup>

### Popper's criticism

Karl Popper approached the problem of indeterminacy as a logician and metaphysical realist.<sup>[61]</sup> He disagreed with the application of the uncertainty relations to individual particles rather than to ensembles of identically prepared particles, referring to them as "statistical scatter relations".<sup>[61][62]</sup> In this statistical interpretation, a *particular* measurement may be made to arbitrary precision without invalidating the quantum theory. This directly contrasts with the Copenhagen interpretation of quantum mechanics, which is non-deterministic but lacks local hidden variables.

In 1934, Popper published *Zur Kritik der Ungenauigkeitsrelationen (Critique of the Uncertainty Relations)* in *Naturwissenschaften*,<sup>[63]</sup> and in the same year *Logik der Forschung* (translated and updated by the author as *The Logic of Scientific Discovery* in 1959), outlining his arguments for the statistical interpretation. In 1982, he further developed his theory in *Quantum theory and the schism in Physics*, writing:

[Heisenberg's] formulae are, beyond all doubt, derivable *statistical formulae* of the quantum theory. But they have been *habitually misinterpreted* by those quantum theorists who said that these formulae can be interpreted as determining some upper limit to the *precision of our measurements*.<sup>[original emphasis]</sup><sup>[64]</sup>

Popper proposed an experiment to falsify the uncertainty relations, although he later withdrew his initial version after discussions with Weizsäcker, Heisenberg, and Einstein; this experiment may have influenced the formulation of the

EPR experiment.<sup>[61][65]</sup>

## Many-worlds uncertainty

The many-worlds interpretation originally outlined by Hugh Everett III in 1957 is partly meant to reconcile the differences between the Einstein and Bohr's views by replacing Bohr's wave function collapse with an ensemble of deterministic and independent universes whose *distribution* is governed by wave functions and the Schrödinger equation. Thus, uncertainty in the many-worlds interpretation follows from each observer within any universe having no knowledge of what goes on in the other universes.

## Free will

Some scientists including Arthur Compton<sup>[66]</sup> and Martin Heisenberg<sup>[67]</sup> have suggested that the uncertainty principle, or at least the general probabilistic nature of quantum mechanics, could be evidence for the two-stage model of free will. The standard view, however, is that apart from the basic role of quantum mechanics as a foundation for chemistry, nontrivial biological mechanisms requiring quantum mechanics are unlikely, due to the rapid decoherence time of quantum systems at room temperature.<sup>[68]</sup>

## Notes

- [1] Paraphrase of Heisenberg's uncertainty paper of 1927.
- [2] Kennard, E. H. (1927), "Zur Quantenmechanik einfacher Bewegungstypen", *Zeitschrift für Physik* **44** (4–5): 326, Bibcode 1927ZPhy...44..326K, doi:10.1007/BF01391200.
- [3] Weyl, H. (1928), *Gruppentheorie Und Quantenmechanik*, Leipzig: Hirzel
- [4] Furuta, Aya (2012), "One Thing Is Certain: Heisenberg's Uncertainty Principle Is Not Dead" (<http://www.scientificamerican.com/article.cfm?id=heisenbergs-uncertainty-principle-is-not-dead>), *Scientific American*,
- [5] Ozawa, Masanao (2003), "Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement", *Physical Review A* **67** (4), arXiv:quant-ph/0207121, Bibcode 2003PhRvA..67d2105O, doi:10.1103/PhysRevA.67.042105
- [6] Werner Heisenberg, *The Physical Principles of the Quantum Theory*, p. 20
- [7] youtube.com website (<http://www.youtube.com/watch?v=TcmGYe39XG0>) Indian Institute of Technology Madras, Professor V. Balakrishnan, Lecture 1 – Introduction to Quantum Physics; Heisenberg's uncertainty principle, National Programme of Technology Enhanced Learning
- [8] Quantum Mechanics Non-Relativistic Theory, Third Edition: Volume 3. Landau, Lifshitz
- [9] Elion, W. J.; M. Matters, U. Geigenmüller & J. E. Mooij (1994), "Direct demonstration of Heisenberg's uncertainty principle in a superconductor", *Nature* **371** (6498): 594–595, Bibcode 1994Natur.371..594E, doi:10.1038/371594a0
- [10] Smithey, D. T.; M. Beck, J. Cooper, M. G. Raymer (1993), "Measurement of number-phase uncertainty relations of optical fields", *Phys. Rev. A* **48** (4): 3159–3167, Bibcode 1993PhRvA..48.3159S, doi:10.1103/PhysRevA.48.3159, PMID 9909968
- [11] Caves, Carlton (1981), "Quantum-mechanical noise in an interferometer", *Phys. Rev. D* **23** (8): 1693–1708, Bibcode 1981PhRvD..23.1693C, doi:10.1103/PhysRevD.23.1693
- [12] *Quantum mechanics*, Wiley-Interscience: Wiley, 1996, pp. 231–233, ISBN 978-0-471-56952-7
- [13] Robertson, H. P. (1929), "The Uncertainty Principle", *Phys. Rev.* **34**: 163–64, Bibcode 1929PhRv...34..163R, doi:10.1103/PhysRev.34.163
- [14] Schrödinger, E. (1930), "Zum Heisenbergschen Unschärfeprinzip", *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse* **14**: 296–303
- [15] Griffiths, David (2005), *Quantum Mechanics*, New Jersey: Pearson
- [16] Riley, K. F.; M. P. Hobson and S. J. Bence (2006), *Mathematical Methods for Physics and Engineering*, Cambridge, pp. 246
- [17] Davidson, E. R. (1965), "On Derivations of the Uncertainty Principle", *J. Chem. Phys.* **42** (4): 1461, Bibcode 1965JChPh..42.1461D, doi:10.1063/1.1696139
- [18] Jackiw, Roman (1968), "Minimum Uncertainty Product, Number-Phase Uncertainty Product, and Coherent States", *J. Math. Phys.* **9** (3): 339, Bibcode 1968JMP....9..339J, doi:10.1063/1.1664585
- [19] Carruthers, P.; Nieto, M. M. (1968), "Phase and Angle Variables in Quantum Mechanics", *Rev. Mod. Phys.* **40** (2): 411, Bibcode 1968RvMP...40..411C, doi:10.1103/RevModPhys.40.411
- [20] L. I. Mandelstam, I. E. Tamm, *The uncertainty relation between energy and time in nonrelativistic quantum mechanics*, 1945 (<http://daarb.narod.ru/mandtamm/index-eng.html>)
- [21] Likharev, K.K.; A.B. Zorin (1985), "Theory of Bloch-Wave Oscillations in Small Josephson Junctions", *J. Low Temp. Phys.* **59** (3/4): 347–382, Bibcode 1985JLTP...59..347L, doi:10.1007/BF00683782
- [22] Anderson, P.W. (1964), "Special Effects in Superconductivity", in Caianiello, E.R., *Lectures on the Many-Body Problem, Vol. 2*, New York: Academic Press

- [23] Steiger, Nathan. "Quantum Uncertainty and Conservation Law Restrictions on Gate Fidelity" (<http://www.physics.byu.edu/Thesis/view.aspx?id=270>). Brigham Young University. . Retrieved 19 June 2011.
- [24] doi: 10.1142/S021773230100576X  
This citation will be automatically completed in the next few minutes. You can jump the queue or expand by hand ([http://en.wikipedia.org/wiki/Template:cite\\_doi/\\_10.1142.2fs021773230100576x?preload=Template:Cite\\_doi/preload&editintro=Template:Cite\\_doi/editintro&action=edit](http://en.wikipedia.org/wiki/Template:cite_doi/_10.1142.2fs021773230100576x?preload=Template:Cite_doi/preload&editintro=Template:Cite_doi/editintro&action=edit))
- [25] Erhart, Jacqueline; Stephan Sponar, Georg Sulyok, Gerald Badurek, Masanao Ozawa, Yuji Hasegawa (2012), "Experimental demonstration of a universally valid error-disturbance uncertainty relation in spin-measurements", *Nature Physics* **8** (3): 185–189, arXiv:1201.1833, Bibcode 2012NatPh...8..185E, doi:10.1038/nphys2194
- [26] Fujikawa, Kazuo (2012), "Universally valid Heisenberg uncertainty relation", *Phys. Rev. A* **85** (6), arXiv:1205.1360, Bibcode 2012PhRvA..85f2117F, doi:10.1103/PhysRevA.85.062117
- [27] Judge, D. (1964), "On the uncertainty relation for angle variables", *Il Nuovo Cimento* **31** (2): 332–340, doi:10.1007/BF02733639
- [28] Bouten, M.; N. Maene, P. Van Leuven (1965), "On an uncertainty relation for angle variables", *Il Nuovo Cimento* **37** (3): 1119–1125, doi:10.1007/BF02773197
- [29] Louisell, W. H. (1963), "Amplitude and phase uncertainty relations", *Physics Letters* **7** (1): 60–61, Bibcode 1963PhL.....7...60L, doi:10.1016/0031-9163(63)90442-6
- [30] DeWitt, B. S.; Graham, N. (1973), *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton: Princeton University Press, pp. 52–53, ISBN 0-691-08126-3
- [31] Hirschman, I. I., Jr. (1957), "A note on entropy", *American Journal of Mathematics* **79** (1): 152–156, doi:10.2307/2372390, JSTOR 2372390.
- [32] Beckner, W. (1975), "Inequalities in Fourier analysis", *Annals of Mathematics* **102** (6): 159–182, doi:10.2307/1970980, JSTOR 1970980.
- [33] Białynicki-Birula, I.; Mycielski, J. (1975), "Uncertainty Relations for Information Entropy in Wave Mechanics", *Communications in Mathematical Physics* **44** (2): 129, Bibcode 1975CMaPh..44..129B, doi:10.1007/BF01608825
- [34] Chafaï, D. (1970), *Gaussian maximum of entropy and reversed log-Sobolev inequality*, pp. 2227, arXiv:math/0102227, Bibcode 2001math.....2227C
- [35] Havin, V.; Jöricke, B. (1994), *The Uncertainty Principle in Harmonic Analysis*, Springer-Verlag
- [36] Folland, Gerald; Sitaram, Alladi (May 1997), "The Uncertainty Principle: A Mathematical Survey" (<http://www.springerlink.com/content/f7492v6374716m4g/fulltext.pdf>) (PDF), *Journal of Fourier Analysis and Applications* **3** (3): 207–238, doi:10.1007/BF02649110, MR98f:42006,
- [37] Sitaram, A (2001), "Uncertainty principle, mathematical" (<http://www.encyclopediaofmath.org/index.php?title=U/u130020>), in Hazewinkel, Michiel, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4,
- [38] Amrein, W.O.; Berthier, A.M. (1977), "On support properties of  $L^p$ -functions and their Fourier transforms", *Journal of Functional Analysis* **24** (3): 258–267, doi:10.1016/0022-1236(77)90056-8.
- [39] Benedicks, M. (1985), "On Fourier transforms of functions supported on sets of finite Lebesgue measure", *J. Math. Anal. Appl.* **106** (1): 180–183, doi:10.1016/0022-247X(85)90140-4
- [40] Nazarov, F. (1994), "Local estimates for exponential polynomials and their applications to inequalities of the uncertainty principle type", *St. Petersburg Math. J.* **5**: 663–717
- [41] Jaming, Ph. (2007), "Nazarov's uncertainty principles in higher dimension", *J. Approx. Theory* **149** (1): 30–41, doi:10.1016/j.jat.2007.04.005
- [42] Hardy, G.H. (1933), "A theorem concerning Fourier transforms", *J. London Math. Soc.* **8** (3): 227–231, doi:10.1112/jlms/s1-8.3.227
- [43] Hörmander, L. (1991), "A uniqueness theorem of Beurling for Fourier transform pairs", *Ark. Mat.* **29**: 231–240, Bibcode 1991ArM....29..237H, doi:10.1007/BF02384339
- [44] Bonami, A.; Demange, B.; Jaming, Ph. (2003), "Hermite functions and uncertainty principles for the Fourier and the windowed Fourier transforms", *Rev. Mat. Iberoamericana* **19**: 23–55., arXiv:math/0102111, Bibcode 2001math.....2111B, doi:10.4171/RMI/337
- [45] *Uncertainty Principles Associated to Non-degenerate Quadratic Forms*, Société Mathématique de France, 2009, ISBN 978-2-85629-297-6
- [46] American Physical Society online exhibit on the Uncertainty Principle (<http://www.aip.org/history/heisenberg/p08.htm>)
- [47] Bohr, Niels (1958), "Atomic Physics and Human Knowledge", *American Journal of Physics* (New York: Wiley) **26** (8): 38, Bibcode 1958AmJPh..26..596B, doi:10.1119/1.1934707
- [48] Heisenberg, W., *Die Physik der Atomkerne*, Taylor & Francis, 1952, p. 30.
- [49] Heisenberg, W. (1927), "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik", *Zeitschrift für Physik* **43** (3–4): 172–198, Bibcode 1927ZPhy...43..172H, doi:10.1007/BF01397280.
- [50] Heisenberg, W. (1930), *Physikalische Prinzipien der Quantentheorie*, Leipzig: Hirzel English translation *The Physical Principles of Quantum Theory*. Chicago: University of Chicago Press, 1930.
- [51] Cassidy, David; Saperstein, Alvin M. (2009), "Beyond Uncertainty: Heisenberg, Quantum Physics, and the Bomb", *Physics Today* (New York: Bellevue Literary Press) **63**: 185, Bibcode 2010PhT....63a..49C, doi:10.1063/1.3293416
- [52] Tipler, Paul A.; Llewellyn, Ralph A. (1999), "5–5", *Modern Physics* (3rd ed.), W. H. Freeman and Co., ISBN 1-57259-164-1
- [53] R&D Magazine & University of Toronto, September 10, 2012 Scientists cast doubt on the uncertainty principle ([http://www.rdmag.com/News/2012/09/General-Science-Scientists-Cast-Doubt-On-The-Uncertainty-Principle/?et\\_cid=2840724&et rid=282345818&linkid=http://www.rdmag.com/News/2012/09/General-Science-Scientists-Cast-Doubt-On-The-Uncertainty-Principle/](http://www.rdmag.com/News/2012/09/General-Science-Scientists-Cast-Doubt-On-The-Uncertainty-Principle/?et_cid=2840724&et rid=282345818&linkid=http://www.rdmag.com/News/2012/09/General-Science-Scientists-Cast-Doubt-On-The-Uncertainty-Principle/)) retrieved Sept 10, 2012
- [54] Feynman lectures on Physics, vol 3, 2–2

- [55] Gamow, G., *The great physicists from Galileo to Einstein*, Courier Dover, 1988, p.260.
- [56] Kumar, M., *Quantum: Einstein, Bohr and the Great Debate About the Nature of Reality*, Icon, 2009, p. 282.
- [57] Gamow, G., *The great physicists from Galileo to Einstein*, Courier Dover, 1988, p. 260–261.
- [58] Kumar, M., *Quantum: Einstein, Bohr and the Great Debate About the Nature of Reality*, Icon, 2009, p. 287.
- [59] Isaacson, Walter (2007), *Einstein: His Life and Universe*, New York: Simon & Schuster, p. 452, ISBN 978-0-7432-6473-0
- [60] Gerardus 't Hooft has at times advocated this point of view.
- [61] Popper, Karl (1959), *The Logic of Scientific Discovery*, Hutchinson & Co.
- [62] Jarvie, Ian Charles; Milford, Karl; Miller, David W (2006), *Karl Popper: a centenary assessment.*, **3**, Ashgate Publishing, ISBN 978-0-7546-5712-5
- [63] Popper, Karl; Carl Friedrich von Weizsäcker (1934), "Zur Kritik der Ungenauigkeitsrelationen (Critique of the Uncertainty Relations)", *Naturwissenschaften* **22** (48): 807–808, Bibcode 1934NW.....22..807P, doi:10.1007/BF01496543.
- [64] Popper, K. *Quantum theory and the schism in Physics*, Unwin Hyman Ltd, 1982, pp. 53–54.
- [65] Mehra, Jagdish; Rechenberg, Helmut (2001), *The Historical Development of Quantum Theory*, Springer, ISBN 978-0-387-95086-0
- [66] A. H. Compton, *Science*, **74**, 1911, August 14, 1931
- [67] Martin Heisenberg, "Is Free Will an Illusion?", *Nature*, **459**, (May 2009): 164–165
- [68] P.C.W. Davies, "Does quantum mechanics play a non-trivial role in life?", *BioSystems*, **78**, (2004): 69–79

## External links

- Hazewinkel, Michiel, ed. (2001), "Uncertainty principle" (<http://www.encyclopediaofmath.org/index.php?title=p/u095100>), *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Annotated pre-publication proof sheet of Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/bond/papers/corr155.1.html>), March 23, 1927.
- Matter as a Wave ([http://www.lightandmatter.com/html\\_books/6mr/ch04/ch04.html](http://www.lightandmatter.com/html_books/6mr/ch04/ch04.html)) – a chapter from an online textbook
- Quantum mechanics: Myths and facts (<http://arxiv.org/abs/quant-ph/0609163>)
- Stanford Encyclopedia of Philosophy entry (<http://plato.stanford.edu/entries/qt-uncertainty/>)
- Fourier Transforms and Uncertainty (<http://www.mathpages.com/home/kmath488/kmath488.htm>) at MathPages
- aip.org: Quantum mechanics 1925–1927 – The uncertainty principle (<http://www.aip.org/history/heisenberg/p08.htm>)
- Eric Weisstein's World of Physics – Uncertainty principle (<http://scienceworld.wolfram.com/physics/UncertaintyPrinciple.html>)
- John Baez on the time-energy uncertainty relation (<http://math.ucr.edu/home/baez/uncertainty.html>)
- The certainty principle (<http://daarb.narod.ru/tcpr-eng.html>)
- Common Interpretation of Heisenberg's Uncertainty Principle Is Proved False (<http://www.scientificamerican.com/article.cfm?id=common-interpretation-of-heisenbergs-uncertainty-principle-is-proven-false>)



# Quantum tunnelling

**Quantum tunnelling** (also spelled **quantum tunneling** in American English) refers to the quantum mechanical phenomenon where a particle tunnels through a barrier that it classically could not surmount. This plays an essential role in several physical phenomena, such as the nuclear fusion that occurs in main sequence stars like the sun,<sup>[1]</sup> and has important applications to modern devices such as the tunnel diode.<sup>[2]</sup> The effect was predicted in the early 20th century and its acceptance, as a general physical phenomenon, came mid-century.<sup>[3]</sup>

Tunnelling is often explained using the Heisenberg uncertainty principle and the wave–particle duality of matter. Purely quantum mechanical concepts are central to the phenomenon, so quantum tunnelling is one of the novel implications of quantum mechanics.

## History

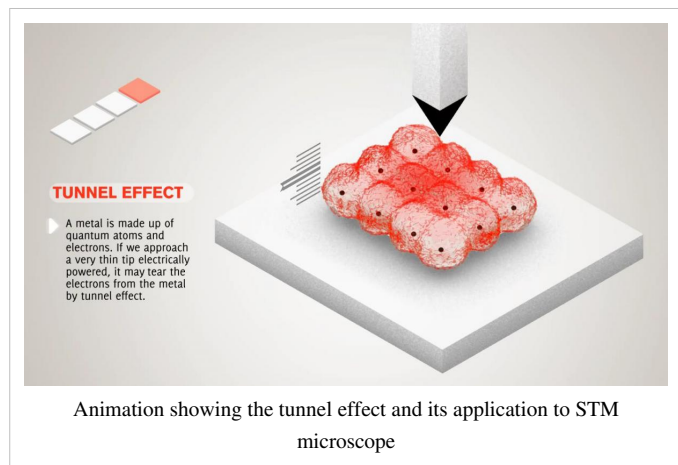
Quantum tunnelling was developed from the study of radioactivity,<sup>[3]</sup> which was discovered in 1896 by Henri Becquerel.<sup>[4]</sup> Radioactivity was examined further by Marie and Pierre Curie, for which they earned the Nobel Prize in Physics in 1903.<sup>[4]</sup> Ernest Rutherford and Egon Schweidler studied its nature, which was later verified empirically by Friedrich Kohlrausch. The idea of the half-life and the impossibility of predicting decay was created from their work.<sup>[3]</sup>

Friedrich Hund was the first to take notice of tunnelling in 1927 when he was calculating the ground state of the double-well potential.<sup>[4]</sup> Its first application was a mathematical explanation for alpha decay, which was done in 1928 by George Gamow and independently by Ronald Gurney and Edward Condon.<sup>[5][6][7]</sup> The two researchers simultaneously solved the Schrödinger equation for a model nuclear potential and derived a relationship between the half-life of the particle and the energy of emission that depended directly on the mathematical probability of tunnelling.

After attending a seminar by Gamow, Max Born recognized the generality of tunnelling. He realized that it was not restricted to nuclear physics, but was a general result of quantum mechanics that applies to many different systems.<sup>[3]</sup> Shortly thereafter, both groups considered the case of particles tunnelling into the nucleus. The study of semiconductors and the development of transistors and diodes led to the acceptance of electron tunnelling in solids by 1957. The work of Leo Esaki, Ivar Giaever and Brian David Josephson predicted the tunnelling of superconducting Cooper pairs, for which they received the Nobel Prize in Physics in 1973.<sup>[3]</sup>

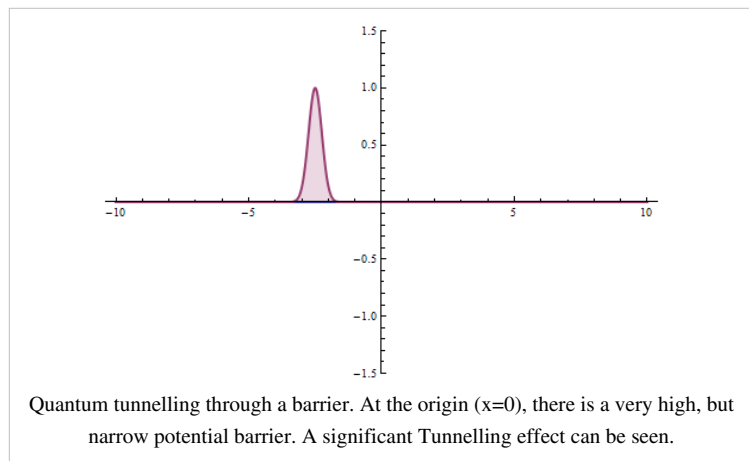
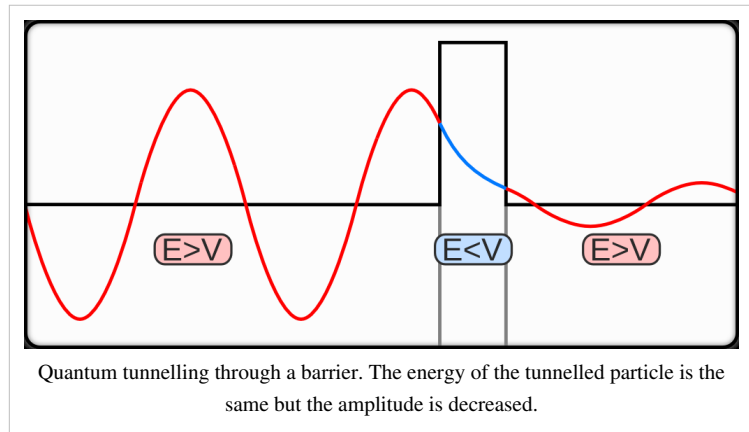
## Introduction to the concept

Quantum tunnelling falls under the domain of quantum mechanics: the study of what happens at the quantum scale. This process cannot be directly perceived, but much of its understanding is shaped by the macroscopic world, which classical mechanics can not adequately explain. To understand the phenomenon, particles attempting to travel between potential barriers can be compared to a ball trying to roll over a hill; quantum mechanics and classical mechanics differ in their treatment of this scenario. Classical



mechanics predicts that particles that do not have enough energy to classically surmount a barrier will not be able to reach the other side. Thus, a ball without sufficient energy to surmount the hill would roll back down. Or, lacking the energy to penetrate a wall, it would bounce back (reflection) or in the extreme case, bury itself inside the wall (absorption). In quantum mechanics, these particles can, with a very small probability, *tunnel* to the other side, thus crossing the barrier. Here, the ball could, in a sense, *borrow* energy from its surroundings to tunnel through the wall or roll over the hill, paying it back by making the reflected electrons more energetic than they otherwise would have been.<sup>[8]</sup>

The reason for this difference comes from the treatment of matter in quantum mechanics as having properties of waves and particles. One interpretation of this duality involves the Heisenberg uncertainty principle, which defines a limit on how precisely the position and the momentum of a particle can be known at the same time.<sup>[4]</sup> This implies that there are no solutions with a probability of exactly zero (or one), though a solution may approach infinity if, for example, the calculation for its position was taken as a probability of 1, the other, i.e. its speed, would have to be infinity. Hence, the probability of a given particle's existence on the opposite side of an intervening barrier is non-zero, and such particles will appear—with no indication of physically transiting the barrier—on the 'other' (a semantically difficult word in this instance) side with a frequency proportional to this probability.



### The tunnelling problem

The wave function of a particle summarizes everything that can be known about a physical system.<sup>[9]</sup> Therefore, problems in quantum mechanics center around the analysis of the wave function for a system. Using mathematical formulations of quantum mechanics, such as the Schrödinger equation, the wave function can be solved. This is directly related to the probability density of the particle's position, which describes the probability that the particle is at any given place. In the limit of large barriers, the probability of tunnelling decreases for taller and wider barriers.

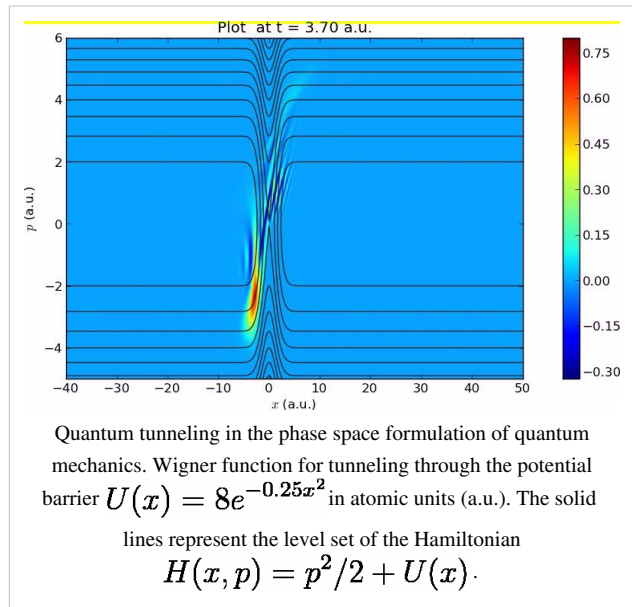
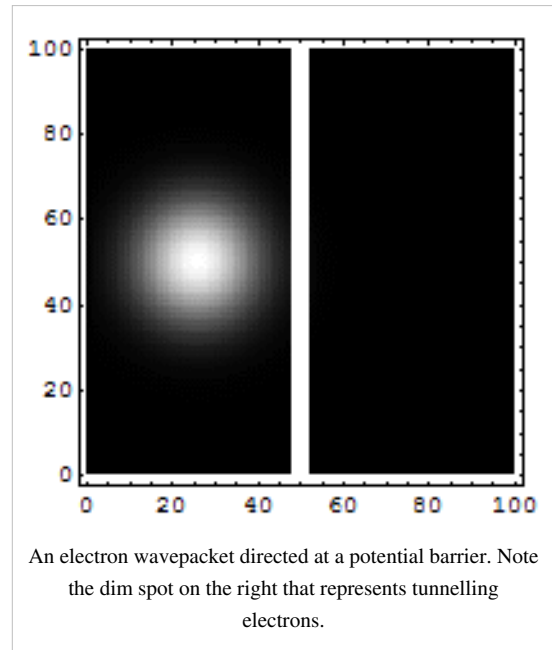
For simple tunnelling-barrier models, such as the rectangular barrier, an analytic solution exists. Problems in real life often do not have one, so "semiclassical" or "quasiclassical" methods have been developed to give approximate solutions to these problems, like the WKB approximation. Probabilities may be derived with arbitrary precision, constrained by computational resources, via Feynman's path integral method; such precision is seldom required in engineering practice.

### Related phenomena

There are several phenomena that have the same behavior as quantum tunnelling, and thus can be accurately described by tunnelling. Examples include the evanescent wave coupling (the application of Maxwell's wave-equation to light) and the application of the non-dispersive wave-equation from acoustics applied to "waves on strings". Evanescent wave coupling, until recently, was only called "tunnelling" in quantum mechanics; now it is used in other contexts.

These effects are modelled similarly to the rectangular potential barrier. In these cases, there is one transmission medium through which the wave propagates that is the same or nearly the same throughout, and a second medium through which the wave travels differently. This can be described as a thin region of medium B between two regions of medium A. The analysis of a rectangular barrier by means of the Schrödinger equation can be adapted to these other effects provided that the wave equation has travelling wave solutions in medium A but real exponential solutions in medium B.

In optics, medium A is a vacuum while medium B is glass. In acoustics, medium A may be a liquid or gas and medium B a solid. For both cases, medium A is a region of space where the particle's total energy is greater than its potential energy and medium B is the potential barrier. These have an incoming wave and resultant waves in both



directions. There can be more mediums and barriers, and the barriers need not be discrete; approximations are useful in this case.

## Applications

Tunnelling occurs with barriers of thickness around 1-3 nm and smaller,<sup>[10]</sup> but is the cause of some important macroscopic physical phenomena. For instance, tunnelling is a source of current leakage in very-large-scale integration (VLSI) electronics and results in the substantial power drain and heating effects that plague high-speed and mobile technology; it is considered the lower limit on how small computer chips can be made.<sup>[11]</sup>

## Radioactive decay

Radioactive decay is the process of emission of particles and energy from the unstable nucleus of an atom to form a stable product. This is done via the tunnelling of a particle out of the nucleus (an electron tunnelling into the nucleus is electron capture). This was the first application of quantum tunnelling and led to the first approximations.

## Spontaneous DNA Mutation

Spontaneous mutation of DNA occurs when normal DNA replication takes place after a particularly significant proton has defied the odds in quantum tunneling in what is called "proton tunneling."<sup>[12]</sup> (quantum bio) A hydrogen bond joins normal base pairs of DNA. There exists a double well potential along a hydrogen bond separated by a potential energy barrier. It is believed that the double well potential is asymmetric with one well deeper than the other so the proton normally rests in the deeper well. For a mutation to occur, the proton must have tunneled into the shallower of the two potential wells. The movement of the proton from its regular position is called a tautomeric transition. If DNA replication takes place in this state, the base pairing rule for DNA may be jeopardized causing a mutation.<sup>[13]</sup> Per-Olov Lowdin was the first to develop this theory of spontaneous mutation within the double helix. (quantum bio) Other instances of quantum tunneling-induced mutations in biology are believed to be the cause of aging and cancer.

## Cold emission

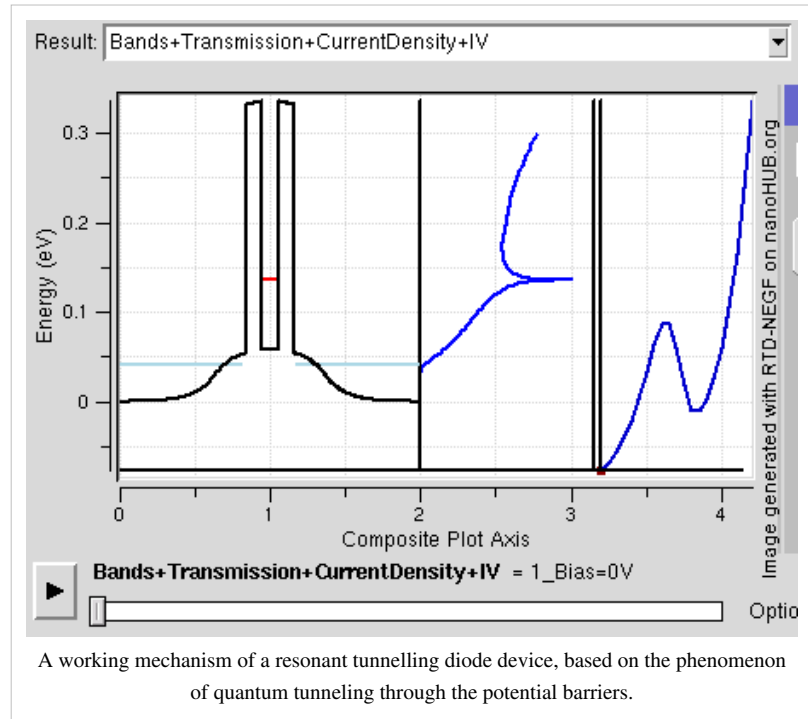
Cold emission of electrons is relevant to semiconductors and superconductor physics. It is similar to thermionic emission, where electrons randomly jump from the surface of a metal to follow a voltage bias because they statistically end up with more energy than the barrier, through random collisions with other particles. When the electric field is very large, the barrier becomes thin enough for electrons to tunnel out of the atomic state, leading to a current that varies approximately exponentially with the electric field.<sup>[14]</sup> These materials are important for flash memory and for some electron microscopes.

## Tunnel junction

A simple barrier can be created by separating two conductors with a very thin insulator. These are tunnel junctions, the study of which requires quantum tunnelling.<sup>[15]</sup> Josephson junctions take advantage of quantum tunnelling and the superconductivity of some semiconductors to create the Josephson effect. This has applications in precision measurements of voltages and magnetic fields,<sup>[14]</sup> as well as the multijunction solar cell.

## Tunnel diode

Diodes are electrical semiconductor devices that allow electric current flow in one direction more than the other. The device depends on a depletion layer between N-type and P-type semiconductors to serve its purpose; when these are very heavily doped the depletion layer can be thin enough for tunnelling. Then, when a small forward bias is applied the current due to tunnelling is significant. This has a maximum at the point where the energy level of the p and n conduction bands are the same. As the voltage bias is increased, the two conduction bands no longer line up and the diode acts typically.<sup>[16]</sup>



Because the tunnelling current drops off rapidly, tunnel diodes can be created that have a range of voltages for which current decreases as voltage is increased. This peculiar property is used in some applications, like high speed devices where the characteristic tunnelling probability changes as rapidly as the bias voltage.<sup>[16]</sup>

The resonant tunnelling diode makes use of quantum tunnelling in a very different manner to achieve a similar result. This diode has a resonant voltage for which there is a lot of current that favors a particular voltage, achieved by placing two very thin layers with a high energy conductance band very near each other. This creates a quantum potential well that have a discrete lowest energy level. When this energy level is higher than that of the electrons, no tunnelling will occur, and the diode is in reverse bias. Once the two voltage energies align, the electrons flow like an open wire. As the voltage is increased further tunnelling becomes improbable and the diode acts like a normal diode again before a second energy level becomes noticeable.<sup>[17]</sup>

## Tunnelling field effect transistor

A European research project has demonstrated field effect transistors in which the gate (channel) is controlled via quantum tunnelling rather than by thermal injection, reducing gate voltage from ~1 volt to 0.2 volts and reducing power consumption by up to 100x. If these transistors can be scaled up into VLSI chips, they will significantly improve the performance per power of integrated circuits.<sup>[18]</sup>

## Quantum conductivity

While the Drude model of electrical conductivity makes excellent predictions about the nature of electrons conducting in metals, it can be furthered by using quantum tunnelling to explain the nature of the electron's collisions.<sup>[14]</sup> When a free electron wave packet encounters a long array of uniformly spaced barriers the reflected part of the wave packet interferes uniformly with the transmitted one between all barriers so that there are cases of 100% transmission. The theory predicts that if positively charged nuclei form a perfectly rectangular array, electrons will tunnel through the metal as free electrons, leading to an extremely high conductance, and that impurities in the metal will disrupt it significantly.<sup>[14]</sup>

## Scanning tunnelling microscope

The scanning tunnelling microscope (STM), invented by Gerd Binnig and Heinrich Rohrer, allows imaging of individual atoms on the surface of a metal.<sup>[14]</sup> It operates by taking advantage of the relationship between quantum tunnelling with distance. When the tip of the STM's needle is brought very close to a conduction surface that has a voltage bias, by measuring the current of electrons that are tunnelling between the needle and the surface, the distance between the needle and the surface can be measured. By using piezoelectric rods that change in size when voltage is applied over them the height of the tip can be adjusted to keep the tunnelling current constant. The time-varying voltages that are applied to these rods can be recorded and used to image the surface of the conductor.<sup>[14]</sup> STMs are accurate to 0.001 nm, or about 1% of atomic diameter.<sup>[17]</sup>

## Faster than light

It is possible for spin zero particles to travel faster than the speed of light when tunnelling.<sup>[3]</sup> This apparently violates the principle of causality (physics), since there will be a frame of reference in which it arrives before it has left. However, careful analysis of the transmission of the wave packet shows that there is actually no violation of relativity theory. In 1998, P.E. Low reviewed briefly the phenomenon of zero time tunneling.<sup>[19]</sup> More recently experimental tunneling time data of phonons, photons, and electrons are published by G. Nimtz.<sup>[20]</sup>

## Mathematical discussions of quantum tunnelling

The following subsections discuss the mathematical formulations of quantum tunnelling.

### The Schrödinger equation

The time-independent Schrödinger equation for one particle in one dimension can be written as

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Psi(x) + V(x)\Psi(x) = E\Psi(x) \text{ or}$$

$$\frac{d^2}{dx^2} \Psi(x) = \frac{2m}{\hbar^2} (V(x) - E) \Psi(x) \equiv \frac{2m}{\hbar^2} M(x)\Psi(x),$$

where  $\hbar$  is the reduced Planck's constant,  $m$  is the particle mass,  $x$  represents distance measured in the direction of motion of the particle,  $\Psi$  is the Schrödinger wave function,  $V$  is the potential energy of the particle (measured relative to any convenient reference level),  $E$  is the energy of the particle that is associated with motion in the  $x$ -axis (measured relative to  $V$ ), and  $M(x)$  is a quantity defined by  $V(x) - E$  which has no accepted name in physics.

The solutions of the Schrödinger equation take different forms for different values of  $x$ , depending on whether  $M(x)$  is positive or negative. When  $M(x)$  is constant and negative, then the Schrödinger equation can be written in the form

$$\frac{d^2}{dx^2}\Psi(x) = \frac{2m}{\hbar^2}M(x)\Psi(x) = -k^2\Psi(x), \quad \text{where } k^2 = -\frac{2m}{\hbar^2}M.$$

The solutions of this equation represent traveling waves, with phase-constant  $+k$  or  $-k$ . Alternatively, if  $M(x)$  is constant and positive, then the Schrödinger equation can be written in the form

$$\frac{d^2}{dx^2}\Psi(x) = \frac{2m}{\hbar^2}M(x)\Psi(x) = \kappa^2\Psi(x), \quad \text{where } \kappa^2 = \frac{2m}{\hbar^2}M.$$

The solutions of this equation are rising and falling exponentials in the form of evanescent waves. When  $M(x)$  varies with position, the same difference in behaviour occurs, depending on whether  $M(x)$  is negative or positive. It follows that the sign of  $M(x)$  determines the nature of the medium, with positive  $M(x)$  corresponding to medium A as described above and negative  $M(x)$  corresponding to medium B. It thus follows that evanescent wave coupling can occur if a region of positive  $M(x)$  is sandwiched between two regions of negative  $M(x)$ , hence creating a potential barrier.

The mathematics of dealing with the situation where  $M(x)$  varies with  $x$  is difficult, except in special cases that usually do not correspond to physical reality. A discussion of the semi-classical approximate method, as found in physics textbooks, is given in the next section. A full and complicated mathematical treatment appears in the 1965 monograph by Fröman and Fröman noted below. Their ideas have not been incorporated into physics textbooks, but their corrections have little quantitative effect.

## The WKB approximation

The wave function is expressed as the exponential of a function:

$$\Psi(x) = e^{\Phi(x)}, \quad \text{where } \Phi''(x) + \Phi'(x)^2 = \frac{2m}{\hbar^2}(V(x) - E).$$

$\Phi'(x)$  is then separated into real and imaginary parts:

$$\Phi'(x) = A(x) + iB(x), \quad \text{where } A(x) \text{ and } B(x) \text{ are real-valued functions.}$$

Substituting the second equation into the first and using the fact that the imaginary part needs to be 0 results in:

$$A'(x) + A(x)^2 - B(x)^2 = \frac{2m}{\hbar^2}(V(x) - E).$$

To solve this equation using the semiclassical approximation, each function must be expanded as a power series in  $\hbar$ . From the equations, the power series must start with at least an order of  $\hbar^{-1}$  to satisfy the real part of the equation; for a good classical limit starting with the highest power of Planck's constant possible is preferable, which leads to

$$A(x) = \frac{1}{\hbar} \sum_{k=0}^{\infty} \hbar^k A_k(x)$$

and

$$B(x) = \frac{1}{\hbar} \sum_{k=0}^{\infty} \hbar^k B_k(x),$$

with the following constraints on the lowest order terms,

$$A_0(x)^2 - B_0(x)^2 = 2m(V(x) - E)$$

and

$$A_0(x)B_0(x) = 0.$$

At this point two extreme cases can be considered.

**Case 1** If the amplitude varies slowly as compared to the phase  $A_0(x) = 0$  and

$$B_0(x) = \pm \sqrt{2m(E - V(x))}$$

which corresponds to classical motion. Resolving the next order of expansion yields

$$\Psi(x) \approx C \frac{e^{i \int dx \sqrt{\frac{2m}{\hbar^2}(E - V(x))} + \theta}}{\sqrt[4]{\frac{2m}{\hbar^2}(E - V(x))}}$$

### Case 2

If the phase varies slowly as compared to the amplitude,  $B_0(x) = 0$  and

$$A_0(x) = \pm \sqrt{2m(V(x) - E)}$$

which corresponds to tunnelling. Resolving the next order of the expansion yields

$$\Psi(x) \approx \frac{C_+ e^{+\int dx \sqrt{\frac{2m}{\hbar^2}(V(x) - E)}} + C_- e^{-\int dx \sqrt{\frac{2m}{\hbar^2}(V(x) - E)}}}{\sqrt[4]{\frac{2m}{\hbar^2}(V(x) - E)}}$$

In both cases it is apparent from the denominator that both these approximate solutions are bad near the classical turning points  $E = V(x)$ . Away from the potential hill, the particle acts similar to a free and oscillating wave; beneath the potential hill, the particle undergoes exponential changes in amplitude. By considering the behaviour at these limits and classical turning points a global solution can be made.

To start, choose a classical turning point,  $x_1$  and expand  $\frac{2m}{\hbar^2}(V(x) - E)$  in a power series about  $x_1$ :

$$\frac{2m}{\hbar^2}(V(x) - E) = v_1(x - x_1) + v_2(x - x_1)^2 + \dots$$

Keeping only the first order term ensures linearity:

$$\frac{2m}{\hbar^2}(V(x) - E) = v_1(x - x_1).$$

Using this approximation, the equation near  $x_1$  becomes a differential equation:

$$\frac{d^2}{dx^2} \Psi(x) = v_1(x - x_1) \Psi(x).$$

This can be solved using Airy functions as solutions.

$$\Psi(x) = C_A Ai(\sqrt[3]{v_1}(x - x_1)) + C_B Bi(\sqrt[3]{v_1}(x - x_1))$$

Taking these solutions for all classical turning points, a global solution can be formed that links the limiting solutions. Given the 2 coefficients on one side of a classical turning point, the 2 coefficients on the other side of a classical turning point can be determined by using this local solution to connect them.

Hence, the Airy function solutions will asymptote into sine, cosine and exponential functions in the proper limits.

The relationships between  $C$ ,  $\theta$  and  $C_+$ ,  $C_-$  are

$$C_+ = \frac{1}{2} C \cos\left(\theta - \frac{\pi}{4}\right)$$

and

$$C_- = -C \sin\left(\theta - \frac{\pi}{4}\right)$$

With the coefficients found, the global solution can be found. Therefore, the transmission coefficient for a particle tunnelling through a single potential barrier is



$$T = \frac{e^{-2 \int_{x_1}^{x_2} dx \sqrt{\frac{2m}{\hbar^2} (V(x) - E)}}}{\left(1 + \frac{1}{4} e^{-2 \int_{x_1}^{x_2} dx \sqrt{\frac{2m}{\hbar^2} (V(x) - E)}\right)^2},$$

where  $x_1, x_2$  are the 2 classical turning points for the potential barrier.

## References

- [1] College Physics Vol. 2 Serway and Vuille
- [2] Taylor, J: *Modern Physics*, page 234. Prentice Hall, 2004.
- [3] Mohsen Razavy, "Quantum Theory of Tunneling", pages 4, 462. World Scientific Publishing Co. 2003
- [4] Nimtz and Haibel, "Zero Time Space", page 1. WILEY-VCH Verlag GmbH & Co. 2008.
- [5] R W Gurney and E U Condon, "Quantum Mechanics and Radioactive Disintegration" *Nature* **122**, 439 (1928); *Phys. Rev* 33, 127 (1929)
- [6] Interview with [[Hans Bethe ([http://www.aip.org/history/ohilist/4504\\_1.html](http://www.aip.org/history/ohilist/4504_1.html))] by Charles Weiner and Jagdish Mehra at Cornell University, 27 October 1966 accessed 5 April 2010]
- [7] Friedlander, Gerhart; Kennedy, Joseph E; Miller, Julian Malcolm (1964). *Nuclear and Radiochemistry, 2nd edition*. New York, London, Sydney: John Wiley & Sons. pp. 225–7. ISBN 978-0-471-86255-0.
- [8] "Quantum Tunneling Time" ([http://cosmos.asu.edu/publications/papers/Quantum Tunelling Time'](http://cosmos.asu.edu/publications/papers/Quantum%20Tunneling%20Time%20AJPO00023.pdf) AJP000023.pdf). *ASU*. Retrieved 2012-01-28.
- [9] Bjorken and Drell, "Relativistic Quantum Mechanics", page 2. Mcgraw-Hill College, 1965.
- [10] Lerner and Trigg, "Encyclopedia of Physics 2nd Ed.", pg 1308, VCH Publishers (1991).
- [11] "Applications of tunneling" ([http://psi.phys.wits.ac.za/teaching/Connell/phys284/2005/lecture-02/lecture\\_02/node13.html](http://psi.phys.wits.ac.za/teaching/Connell/phys284/2005/lecture-02/lecture_02/node13.html)). Simon Connell 2006.
- [12] Matta, Cherif F. *Quantum Biochemistry: Electronic Structure and Biological Activity*. Weinheim, Germany: Wiley-Vch Verlag GmbH & Co., 2010. eBook. <[http://books.google.com/books?id=a4JhVFauOjgC&printsec=frontcover&source=gbs\\_ge\\_summary\\_r&cad=0](http://books.google.com/books?id=a4JhVFauOjgC&printsec=frontcover&source=gbs_ge_summary_r&cad=0)
- [13] Majumdar, Rabi. *Quantum Mechanics: In Physics and Chemistry with Applications to Bioloty*. New Delhi: PHI Learning Pvt. Ltd., 2011. eBook. <[http://books.google.com/books?id=IJDvyNVeBiYC&printsec=frontcover&source=gbs\\_ge\\_summary\\_r&cad=0](http://books.google.com/books?id=IJDvyNVeBiYC&printsec=frontcover&source=gbs_ge_summary_r&cad=0)
- [14] Taylor, J: *Modern Physics*, page 479. Prentice Hall, 2004.
- [15] Lerner and Trigg, "Encyclopedia of Physics 2nd Ed.", pg 1308-1309. VCH Publishers, 1991.
- [16] Kenneth Krane, "Modern Physics", pg 423. John Wiley and Sons, 1983.
- [17] R. D. Knight, "Physics for Scientists and Engineers: With Modern Physics", pg 1311. Pearson Education, 2004.
- [18] Tunnel field-effect transistors as energy-efficient electronic switches (<http://www.nature.com/nature/journal/v479/n7373/full/nature10679.html>), Adrian M. Ionescu and Heike Riel, *Nature* 479, pp. 329–337 (17 November 2011); doi:10.1038/nature10679
- [19] F.E. Low, Comments on apparent superluminal propagation, *Ann. Phys. (Leipzig)*, 7, 660-661, (1998)
- [20] G. Nimtz, Tunneling confronts special relativity, *Found.Phys.*, 41, 1193, (2011)

## Further reading

- N. Fröman and P.-O. Fröman (1965). *JWKB Approximation: Contributions to the Theory*. Amsterdam: North-Holland.
- Razavy, Mohsen (2003). *Quantum Theory of Tunneling*. World Scientific. ISBN 981-238-019-1.
- Griffiths, David J. (2004). *Introduction to Quantum Mechanics* (2nd ed.). Prentice Hall. ISBN 0-13-805326-X.
- James Binney and Skinner, D. (2010). *The Physics of Quantum Mechanics: An Introduction* (3rd ed.). Cappelra Archive. ISBN 1-902918-51-7.
- Liboff, Richard L. (2002). *Introductory Quantum Mechanics*. Addison-Wesley. ISBN 0-8053-8714-5.
- Vilenkin, Alexander; Vilenkin, Alexander; Winitzki, Serge (2003). "Particle creation in a tunneling universe". *Physical Review D* **68** (2): 023520. arXiv:gr-qc/0210034. Bibcode 2003PhRvD..68b3520H. doi:10.1103/PhysRevD.68.023520.

## External links

- Animation, applications and research linked to tunnel effect and other quantum phenomena (<http://www.toutestquantique.fr/#tunnel>) (Université Paris Sud)
- Animated illustration of quantum tunnelling (<http://molecularmodelingbasics.blogspot.com/2009/09/tunneling-and-stm.html>)
- Animated illustration of quantum tunnelling in a RTD device (<http://nanohub.org/resources/8799>)

# Implicate and explicate order according to David Bohm

---

According to David Bohm's theory, **implicate and explicate orders** are characterised by:

In the enfolded [or implicate] order, space and time are no longer the dominant factors determining the relationships of dependence or independence of different elements. Rather, an entirely different sort of basic connection of elements is possible, from which our ordinary notions of space and time, along with those of separately existent material particles, are abstracted as forms derived from the deeper order. These ordinary notions in fact appear in what is called the "explicate" or "unfolded" order, which is a special and distinguished form contained within the general totality of all the implicate orders (Bohm 1980, p. xv).

## David Bohm's challenges to some generally prevailing views

In proposing this new notion of order, David Bohm explicitly challenged a number of tenets that he believed are fundamental to much scientific work. Bohm challenges the ideas:

1. that phenomena are reducible to fundamental particles and laws describing the behaviour of particles, or more generally to any static (i.e., unchanging) entities, whether separate events in space-time, quantum states, or static entities of some other nature;
  2. related to (1), that human knowledge is *most fundamentally* concerned with mathematical prediction of statistical aggregates of particles;
  3. that an analysis or description of any aspect of reality (e.g., quantum theory, the speed of light) can be unlimited in its domain of relevance;
  4. that the Cartesian coordinate system, or its extension to a curvilinear system, is the deepest conception of underlying order as a basis for analysis and description of the world;
  5. that there is ultimately a sustainable *distinction* between reality and thought, and that there is a corresponding distinction between the observer and observed in an experiment or any other situation (other than a distinction between relatively separate entities valid in the sense of explicate order); and
  6. that it is, in principle, possible to formulate a final notion concerning the nature of reality, i.e., a Theory of Everything.
-

Bohm's proposals have at times been dismissed largely on the basis of such tenets, without due consideration necessarily given to the fact that they had been challenged by Bohm.

Bohm's paradigm is inherently antithetical to reductionism in most forms and accordingly can be regarded as a form of ontological holism. On this, Bohm noted of prevailing views among physicists that "the world is assumed to be constituted of a set of separately existent, indivisible, and unchangeable 'elementary particles', which are the fundamental 'building blocks' of the entire universe ... there seems to be an unshakable faith among physicists that either such particles, or some other kind yet to be discovered, will eventually make possible a complete and coherent explanation of everything" (Bohm 1980, p. 173).

In Bohm's conception of order, then, primacy is given to the undivided whole, and the implicate order inherent within the whole, rather than to parts of the whole, such as particles, quantum states, and continua. For Bohm, the whole encompasses all things, structures, abstractions, and processes, including processes that result in (relatively) stable structures as well as those that involve a metamorphosis of structures or things. In this view, parts may be entities normally regarded as physical, such as atoms or subatomic particles, but they may also be abstract entities, such as quantum states. Whatever their nature and character, according to Bohm, these parts are considered in terms of the whole, and in such terms, they constitute relatively autonomous and independent "sub-totalities." The implication of the view is, therefore, that nothing is *fundamentally* separate or autonomous.

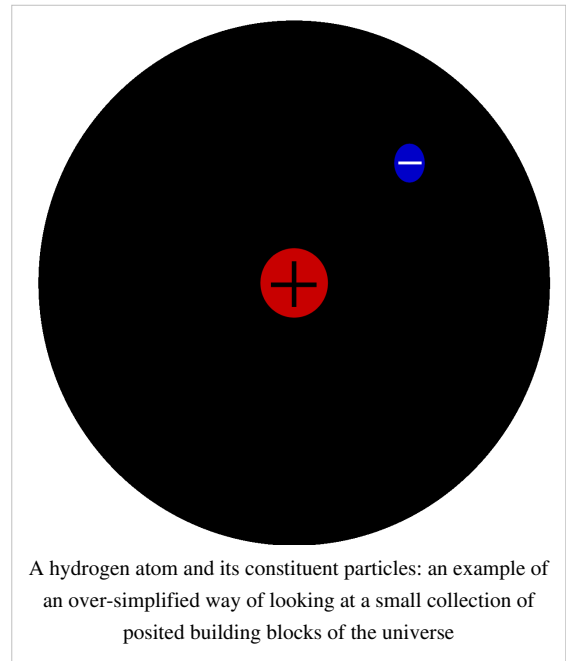
Bohm 1980, p. 11, said: "The new form of insight can perhaps best be called Undivided Wholeness in Flowing Movement. This view implies that flow is in some sense prior to that of the 'things' that can be seen to form and dissolve in this flow." According to Bohm, a vivid image of this sense of analysis of the whole is afforded by vortex structures in a flowing stream. Such vortices can be relatively stable patterns within a continuous flow, but such an analysis does not imply that the flow patterns have any sharp division, or that they are literally separate and independently existent entities; rather, they are most fundamentally undivided. Thus, according to Bohm's view, the whole is in continuous flux, and hence is referred to as the holomovement (movement of the whole).

### Quantum theory and relativity theory

A key motivation for Bohm in proposing a new notion of order was the well-known incompatibility of quantum theory with relativity theory. Bohm 1980, p. xv summarised the state of affairs he perceived to exist:

...in relativity, movement is continuous, causally determinate and well defined, while in quantum mechanics it is discontinuous, not causally determinate and not well-defined. Each theory is committed to its own notions of essentially static and fragmentary modes of existence (relativity to that of separate events connectible by signals, and quantum mechanics to a well-defined quantum state). One thus sees that a new kind of theory is needed which drops these basic commitments and at most recovers some essential features of the older theories as abstract forms derived from a deeper reality in which what prevails is unbroken wholeness.

Bohm maintained that relativity and quantum theories are in basic contradiction in these essential respects, and that a new concept of order should begin with that toward which both theories point: undivided wholeness. This should not be taken to mean that he advocated such powerful theories be discarded. He argued that each was relevant in a



certain context—i.e., a set of interrelated conditions within the explicate order—rather than having unlimited scope, and that apparent contradictions stem from attempts to overgeneralize by superposing the theories on one another, implying greater generality or broader relevance than is ultimately warranted. Thus, Bohm 1980, pp. 156–167 argued: "... in sufficiently broad contexts such analytic descriptions cease to be adequate ... 'the law of the whole' will generally include the possibility of describing the 'loosening' of aspects from each other, so that they will be relatively autonomous in limited contexts ... however, any form of relative autonomy (and heteronomy) is ultimately limited by holonomy, so that in a broad enough context such forms are seen to be merely aspects, relevated in the holomovement, rather than disjoint and separately existent things in interaction."

### Hidden variable theory

Bohm proposed a hidden variable theory of quantum physics (see Bohm interpretation). According to Bohm, a key motivation for doing so was purely to show the possibility of such theories. On this, Bohm 1980, p. 81 said, "... it should be kept in mind that before this proposal was made there had existed the widespread impression that no conception of any hidden variable at all, not even if it were abstract and hypothetical, could possibly be consistent with the quantum theory." Bohm 1980, p. 110 also claimed that "the demonstration of the possibility of theories of hidden variables may serve in a more general philosophical sense to remind us of the unreliability of conclusions based on the assumption of the complete universality of certain features of a given theory, however general their domain of validity seems to be." Another aspect of Bohm's motivation was to point out a confusion he perceived to exist in quantum theory. On the dominant approaches in quantum theory, he said: "...we wish merely to point out that this whole line of approach re-establishes at the abstract level of statistical potentialities the same kind of analysis into separate and autonomous components in interaction that is denied at the more concrete level of individual objects" (Bohm 1980, p. 174).

### The implicate order as an algebra

David Bohm, his co-worker Basil Hiley, and other physicists of Birkbeck college, University of London, worked toward representing the implicate order in form of an appropriate algebra or other pregeometry. They considered spacetime itself as part of an explicit order that is connected to an implicit order that they called *pre-space*. The spacetime manifold and properties of locality and nonlocality then arise from an order in such pre-space. A. M. Frescura and Hiley suggested that an implicate order could be carried by an algebra, with the explicate order being contained in the various representations of this algebra.<sup>[1]</sup> (See also: Work by Bohm and Hiley on implicate orders, pre-space and algebraic structures)

In analogy to Alfred Whitehead's notion of *actual occurrence*, Bohm considered the notion of *moment*—a moment being a not entirely localizable event, with events being allowed to overlap<sup>[2]</sup> and being connected in an over-all implicate order:<sup>[3]</sup>

I propose that each moment of time is a projection from the total implicate order. The term *projection* is a particularly happy choice here, not only because its common meaning is suitable for what is needed, but also because its mathematical meaning as a projection operation,  $P$ , is just what is required for working out these notions in terms of the quantum theory.

Bohm emphasized the primary role of the implicate order's structure:<sup>[4]</sup>

My attitude is that the mathematics of the quantum theory deals *primarily* with the structure of the implicate pre-space and with how an explicate order of space and time emerges from it, rather than with movements of physical entities, such as particles and fields. (This is a kind of extension of what is done in general relativity, which deals primarily with geometry and only secondarily with the entities that are described within this geometry.)

## Quantum entanglement

Central to Bohm's schema are correlations between observables of entities which seem separated by great distances in the explicate order (such as a particular electron here on earth and an alpha particle in one of the stars in the Abell 1835 galaxy, the farthest galaxy from Earth known to humans), manifestations of the implicate order. Within quantum theory there is entanglement of such objects.

This view of order necessarily departs from any notion which entails signalling, and therefore causality. The correlation of observables does not imply a causal influence, and in Bohm's schema the latter represents 'relatively' independent events in space-time; and therefore explicate order.

## Ink droplet analogy

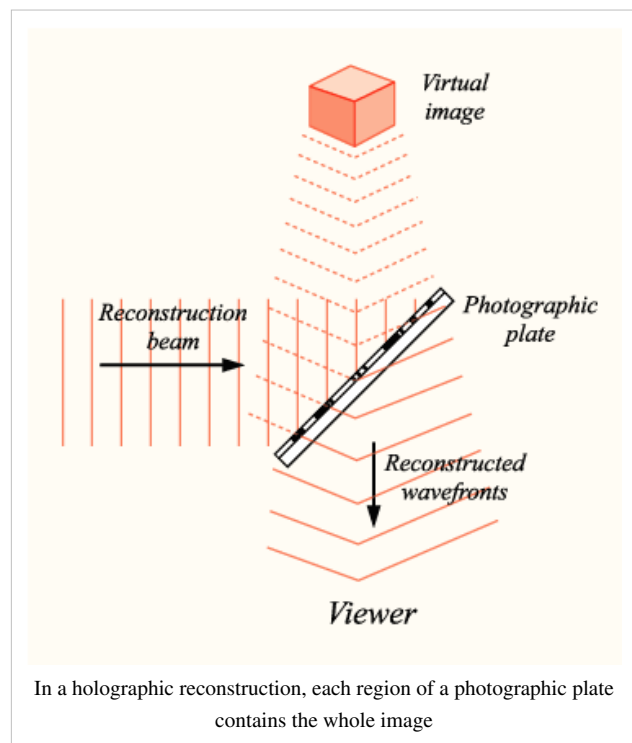
Bohm also used the term *unfoldment* to characterise processes in which the explicate order becomes relevant (or "relevated"). Bohm likens unfoldment also to the decoding of a television signal to produce a sensible image on a screen. The signal, screen, and television electronics in this analogy represent the implicate order whilst the image produced represents the explicate order. He also uses an example in which an ink droplet can be introduced into a highly viscous substance (such as glycerine), and the substance rotated very slowly such that there is negligible diffusion of the substance. In this example, the droplet becomes a thread which, in turn, eventually becomes invisible. However, by rotating the substance in the reverse direction, the droplet can essentially reform. When it is invisible, according to Bohm, the order of the ink droplet as a pattern can be said to be *implicate* within the substance.

In another analogy, Bohm asks us to consider a pattern produced by making small cuts in a folded piece of paper and then, literally, unfolding it. Widely separated elements of the pattern are, in actuality, produced by the same original cut in the folded piece of paper. Here the cuts in the folded paper represent the implicate order and the unfolded pattern represents the explicate order.

## The hologram as analogy for the implicate order

Bohm employed the hologram as a means of characterising implicate order, noting that each region of a photographic plate in which a hologram is observable contains within it the whole three-dimensional image, which can be viewed from a range of perspectives. That is, each region contains a whole and undivided image. In Bohm's words:

There is the germ of a new notion of order here. This order is not to be understood solely in terms of a regular arrangement of objects (e.g., in rows) or as a regular arrangement of events (e.g. in a series). Rather, a total order is contained, in some implicit sense, in each region of space and time. Now, the word 'implicit' is based on the verb 'to implicate'. This means 'to fold inward' ... so we may be led to explore the notion that in some sense each region



contains a total structure 'enfolded' within it".<sup>[5]</sup>

Bohm noted that although the hologram conveys undivided wholeness, it is nevertheless static.

In this view of order, laws represent invariant relationships between explicate entities and structures, and thus Bohm maintained that in physics, the explicate order generally reveals itself within well-constructed experimental contexts as, for example, in the sensibly observable results of instruments. With respect to implicate order, however, Bohm asked us to consider the possibility instead "that physical law should refer primarily to an order of undivided wholeness of the content of description similar to that indicated by the hologram rather than to an order of analysis of such content into separate parts ...".<sup>[6]</sup>

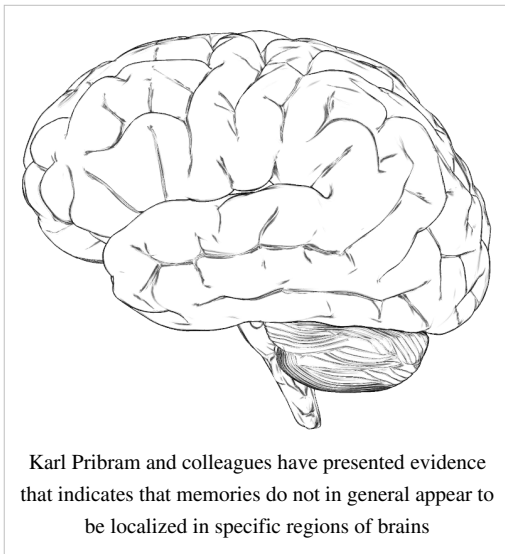
## Implicate orders in art

In the work *Science, Order, and Creativity* (Bohm and Peat, 1987), examples of implicate orders in science are laid out, as well as implicate orders which relate to painting, poetry, and music.

Bohm and Peat emphasize the role of orders of varying complexity, which influence the perception of a work of art as a whole. They note that implicate orders are accessible to human experience. They refer for instance to earlier notes which reverberate when listening to music, or various resonances of words and images which are perceived when reading or hearing poetry.

Christopher Alexander discussed his work in person with Bohm, and pointed out connections among his work and Bohm's notion of an implicate order in *The Nature of Order*.<sup>[7]</sup>

## A common grounding for consciousness and matter



Karl Pribram and colleagues have presented evidence that indicates that memories do not in general appear to be localized in specific regions of brains

The implicate order represents the proposal of a general metaphysical concept in terms of which it is claimed that matter and consciousness might both be understood, in the sense that it is proposed that both matter and consciousness: (i) enfold the structure of the whole within each region, and (ii) involve continuous processes of enfoldment and unfoldment. For example, in the case of matter, entities such as atoms may represent continuous enfoldment and unfoldment which manifests as a relatively stable and autonomous entity that can be observed to follow a relatively well-defined path in space-time. In the case of consciousness, Bohm pointed toward evidence presented by Karl Pribram that memories may be enfolded within every region of the brain rather than being localized (for example in particular regions of the brain, cells, or atoms).

Bohm went on to say:

As in our discussion of matter in general, it is now necessary to go into the question of how in consciousness the explicate order is what is manifest ... the manifest content of consciousness is based essentially on memory, which is what allows such content to be held in a fairly constant form. Of course, to make possible such constancy it is also necessary that this content be organized, not only through relatively fixed association but also with the aid of the rules of logic, and of our basic categories of space, time causality, universality, etc. ... there will be a strong background of recurrent stable, and separable features, against which the transitory and changing aspects of the unbroken flow of experience will be seen as fleeting impressions that tend to be arranged and ordered mainly in terms of the vast totality of the relatively static and fragmented content of [memories].<sup>[8]</sup>

Bohm also claimed that "as with consciousness, each moment has a certain explicate order, and in addition it enfolds all the others, though in its own way. So the relationship of each moment in the whole to all the others is implied by its total content: the way in which it 'holds' all the others enfolded within it". Bohm characterises consciousness as a process in which at each moment, content that was previously implicate is presently explicate, and content which was previously explicate has become implicate.

One may indeed say that our memory is a special case of the process described above, for all that is recorded is held enfolded within the brain cells and these are part of matter in general. The recurrence and stability of our own memory as a relatively independent sub-totality is thus brought about as part of the very same process that sustains the recurrence and stability in the manifest order of matter in general. It follows, then, that the explicate and manifest order of consciousness is not ultimately distinct from that of matter in general.<sup>[9]</sup>

## Notes

- [1] F. A. M. Frescura, B. J. Hiley: Algebras, quantum theory and pre-space (<http://www.bbk.ac.uk/tpru/BasilHiley/P12FrescandHiley3.pdf>), p. 3–4 (published in *Revista Brasileira de Fisica*, Volume Especial, Julho 1984, Os 70 anos de Mario Schonberg, pp. 49-86)
- [2] David Bohm: *Time, the implicate order, and pre-space*, In: David R. Griffin: *Physics and the Ultimate Significance of Time*, State University of New York Press, 1986, ISBN 0-88706-113-3, pp. 177–208, p. 183 ([http://books.google.de/books?id=hXWKzPFgv\\_wC&pg=PA183](http://books.google.de/books?id=hXWKzPFgv_wC&pg=PA183))
- [3] David Bohm: *Time, the implicate order, and pre-space*, In: David R. Griffin: *Physics and the Ultimate Significance of Time*, State University of New York Press, 1986, ISBN 0-88706-113-3, pp. 177–208, p. 189 ([http://books.google.de/books?id=hXWKzPFgv\\_wC&pg=PA189](http://books.google.de/books?id=hXWKzPFgv_wC&pg=PA189))
- [4] David Bohm: *Time, the implicate order, and pre-space*, In: David R. Griffin: *Physics and the Ultimate Significance of Time*, State University of New York Press, 1986, ISBN 0-88706-113-3, pp. 177–208, p. 192–193 ([http://books.google.de/books?id=hXWKzPFgv\\_wC&pg=PA192](http://books.google.de/books?id=hXWKzPFgv_wC&pg=PA192))
- [5] Bohm 1980, p. 149
- [6] Bohm 1980, p. 147
- [7] Christopher Alexander: *The Nature of Order, Book 4 – The Luminous Ground: An Essay on the Art of Building and the Nature of the Universe*, Center for Environmental Structure, ISBN 978-0-9726529-4-0, Footnotes 19 and 20 on p. 336 ([http://books.google.com/books?id=6CIHB3\\_1tLMC&pg=PA336](http://books.google.com/books?id=6CIHB3_1tLMC&pg=PA336)), cited on p. 323 ([http://books.google.com/books?id=6CIHB3\\_1tLMC&pg=PA323](http://books.google.com/books?id=6CIHB3_1tLMC&pg=PA323))
- [8] Bohm 1980, p. 205
- [9] Bohm 1980, p. 208

## References

- Bohm, David (1980), *Wholeness and the Implicate Order*, London: Routledge, ISBN 0-7100-0971-2
- Bohm, David; Hiley, B. J. (1993), *The Undivided Universe*, London: Routledge, ISBN 0-415-06588-7
- Kauffman, S. (1995). *At Home in the Universe*. New York: Oxford University Press. hardcover: ISBN 0-19-509599-5, paperback ISBN 0-19-511130-3
- Kauffman, S. (2000). *Investigations*. New York: Oxford University Press.
- Kuhn, T.S. (1961). *The function of measurement in modern physical science*. *ISIS*, 52, 161-193.
- Schopenhauer, A. (1819/1995) (1995), *The World as Will and Idea*, London: Everyman, ISBN 0-460-87505-1 (D. Derman, Ed.; J. Berman, Trans.).

## Further reading

- Michael Talbot. *The Holographic Universe*, Harpercollins (1991)
- Paavo Pylkkänen. *Cognition, the implicate order and rainforest realism* ([https://helda.helsinki.fi/bitstream/handle/10138/37374/Futura\\_2\\_2012\\_Pylkkanen.pdf?sequence=2](https://helda.helsinki.fi/bitstream/handle/10138/37374/Futura_2_2012_Pylkkanen.pdf?sequence=2)), Futura, vol. 31, no. 2/2012, pp. 74–83.

## External links

- Interview with David Bohm (<http://www.f davidpeat.com/interviews/bohm.htm>) – An interview with Bohm concerning this particular subject matter conducted by F. David Peat.
- Excerpt from *The Holographic Universe* (<http://web.archive.org/web/20041207182354/www.soultravel.nu/2004/040907-swedenborg/index.shtml>) – Parallels some of the experiences of 18th century Swedish mystic, Emanuel Swedenborg, with David Bohm's ideas.
- Thought Knowledge Perception Institute Implicate Order Page (<http://tkpi.org/tags/implicate-order>)

# Biocentrism (theory of everything)

---

**Biocentrism** (from Greek: βίος, *bios*, "life"; and κέντρον, *kentron*, "center") — also known as the **biocentric universe** — is a concept proposed in 2007 by American Doctor of medicine Robert Lanza, a scientist in the fields of biology and regenerative medicine.<sup>[1][2][3]</sup> The theory espouses that biology is the central driving science in the universe, and sees an understanding of the other sciences as reliant on a deeper understanding of biology. The theory states that life and biology are central to being, reality, and the cosmos — life creates the universe rather than the other way around. Biocentrism asserts that current theories of the physical world do not work, and can never be made to work, until they fully account for life and consciousness. While physics is considered fundamental to the study of the universe, and chemistry fundamental to the study of life, biocentrism tries to place biology before the other sciences to produce a *theory of everything*.<sup>[4][5][6]</sup>

Critics have questioned whether the theory is falsifiable. Lanza has claimed that future experiments, such as scaled-up quantum superposition, will either support or contradict the theory.<sup>[7]</sup>

## Hypothesis

Biocentrism was first proposed in a 2007 article by Robert Lanza that appeared in *The American Scholar*, where the goal was to show how biology could build upon quantum physics.<sup>[6]</sup> Two years later, Lanza published a book with astronomer and author Bob Berman entitled *Biocentrism: How Life and Consciousness Are the Keys to Understanding the True Nature of the Universe*, which expanded upon the ideas that Lanza wrote about in his essay for the *Scholar*. The book was featured in news sources including the *San Francisco Chronicle*, The Huffington Post, and MSNBC.<sup>[8][9][10]</sup>

Biocentrism argues that the primacy of consciousness features in the work of Descartes, Kant, Leibniz, Berkeley, Schopenhauer, and Bergson.<sup>[6]</sup> He sees this as supporting the central claim that what we call space and time are forms of animal sense perception, rather than external physical objects.<sup>[9][11]</sup> Lanza argues that biocentrism offers insight into several major puzzles of science, including Heisenberg's uncertainty principle, the double-slit experiment, and the fine tuning of the forces, constants, and laws that shape the universe as we perceive it.<sup>[12]</sup> According to a *Discover* magazine article adapted from Lanza's book, "biocentrism offers a more promising way to bring together all of physics, as scientists have been trying to do since Einstein's unsuccessful unified field theories of eight decades ago."<sup>[13]</sup>

---



## Synopsis of Lanza's book *Biocentrism*

According to Lanza's book, *Biocentrism* suggests that life is not an accidental byproduct of physics, but rather is a key part of our understanding of the universe.<sup>[14]</sup> The theory states that there is no independent external universe outside of biological existence.<sup>[15]</sup> Part of what the theory sees as evidence of this is that there are over 200 physical parameters within the universe so exact that it is seen as more probable that they are that way in order to allow for existence of life and consciousness, rather than coming about at random.<sup>[16]</sup> Biocentrism claims that allowing the observer into the equation opens new approaches to understanding cognition. Through this, biocentrism purports to offer a way to unify the laws of the universe.<sup>[13]</sup>

## Reception

Some physicists have commented that biocentrism currently does not make testable predictions.<sup>[17]</sup> Dr. Vinod Kumar Wadhawan and Ajita Kamal responded to the idea by stating that "The biocentrism approach does not provide any new information about the nature of consciousness, and relies on ignoring recent advances in understanding consciousness from a scientific perspective."<sup>[11]</sup> Biologist PZ Meyers supported their views as well.<sup>[18]</sup> Arizona State University physicist Lawrence Krauss stated, "It may represent interesting philosophy, but it doesn't look, at first glance, as if it will change anything about science."<sup>[17]</sup> In *USA Today Online*, theoretical physicist and science writer David Lindley asserted that Lanza's concept was a "vague, inarticulate metaphor" and stated that "I certainly don't see how thinking his way would lead you into any new sort of scientific or philosophical insight. That's all very nice, I would say to Lanza, but now what? I [also] take issue with his views about physics."<sup>[19]</sup> Richard Conn Henry, Professor of Physics and Astronomy at Johns Hopkins University writing for the *Journal of Scientific Exploration* said: "What Lanza says in this book is not new. Then why does Robert have to say it at all? It is because we, the physicists, do NOT say it—or if we do say it, we only whisper it, and in private—furiously blushing as we mouth the words. True, yes; politically correct, hell no!"<sup>[20]</sup>

Nobel laureate in Physiology or Medicine and physician E. Donnall Thomas said of biocentrism, "Any short statement does not do justice to such a scholarly work. The work is a scholarly consideration of science and philosophy that brings biology into the central role in unifying the whole."<sup>[17]</sup> Indian physician and writer<sup>[21]</sup> Deepak Chopra stated that "Lanza's insights into the nature of consciousness [are] original and exciting" and that "his theory of biocentrism is consistent with the most ancient wisdom traditions of the world which says that consciousness conceives, governs, and becomes a physical world. It is the ground of our Being in which both subjective and objective reality come into existence."<sup>[22]</sup> Daniel Dennett said that he does not believe that the idea met the criteria of a theory in philosophy.<sup>[17]</sup>

## References

- [1] Dr. Robert Lanza | Gary Rabin | biotechnology | Senior Executive Officers | Advanced Cell Technology (<http://www.advancedcell.com/company/leadership-team/senior-executive-officers/>)
- [2] Anthony Atala, Robert Lanza, James A. Thomson, Robert Nerem (2012). *Principles of Regenerative Medicine* ([http://books.google.ca/books?id=Wd\\_wNMFBDxQC&printsec=frontcover&dq=biocentrism+lanza&source=bl&ots=m2RKwIzgL3&sig=MFaWNM6HbUIrfBZ5jff8C9pen1Y&hl=en&sa=X&ei=T58\\_UKitHurz0gGg84DwDA&sqi=2&redir\\_esc=y#v=onepage&q&f=false](http://books.google.ca/books?id=Wd_wNMFBDxQC&printsec=frontcover&dq=biocentrism+lanza&source=bl&ots=m2RKwIzgL3&sig=MFaWNM6HbUIrfBZ5jff8C9pen1Y&hl=en&sa=X&ei=T58_UKitHurz0gGg84DwDA&sqi=2&redir_esc=y#v=onepage&q&f=false)). Academic Press. . Retrieved August 30, 2012.
- [3] Robert Lanza, Irina Klimanskaya (2008). *Essential Stem Cell Methods* ([http://books.google.ca/books?id=aPcvphPV6AgC&printsec=frontcover&dq=biocentrism+lanza&source=bl&ots=e1ZJmkTt5Q&sig=rrG5u34PWCT1VZZUZCSMemMPLfY&hl=en&sa=X&ei=T58\\_UKitHurz0gGg84DwDA&sqi=2&redir\\_esc=y#v=onepage&q&f=false](http://books.google.ca/books?id=aPcvphPV6AgC&printsec=frontcover&dq=biocentrism+lanza&source=bl&ots=e1ZJmkTt5Q&sig=rrG5u34PWCT1VZZUZCSMemMPLfY&hl=en&sa=X&ei=T58_UKitHurz0gGg84DwDA&sqi=2&redir_esc=y#v=onepage&q&f=false)). Academic Press. . Retrieved August 30, 2012.
- [4] Aaron Rowe (2007-03-08). "Will Biology Solve the Universe?" (<http://www.wired.com/medtech/genetics/news/2007/03/72910>). *Wired*. . Retrieved 2009-03-15.
- [5] "Biocentrism" (<http://www.the-scientist.com/news/display/55621/>). *The Scientist*. . Retrieved 2009-04-17.
- [6] "A New Theory of the Universe" (<http://www.theamericanscholar.org/a-new-theory-of-the-universe/>), Spring 2007 *The American Scholar*
- [7] Eric Berger (2009-08-23). "Book Spotlight: Biocentrism" ([http://blogs.chron.com/sciguy/archives/2009/08/book\\_spotlight\\_biocentrism.html](http://blogs.chron.com/sciguy/archives/2009/08/book_spotlight_biocentrism.html)). *Houston Chronicle Blogs*. . Retrieved 2009-12-10.

- [8] Robert Lanza and Deepak Chopra (October 19, 2009). "The illusion of past, present, future" (<http://www.sfgate.com/opinion/chopra/article/The-illusion-of-past-present-future-2473758.php#ixzz253bhvvnv>). *San Francisco Chronicle*. . Retrieved August 30, 2012.
- [9] "Biocentrism: How life creates the universe" (<http://www.msnbc.msn.com/id/31393080/>), *MSNBC.com*
- [10] Robert Lanza (August 12, 2012). "Does Death Exist? New Theory Says 'No'" ([http://www.huffingtonpost.com/robert-lanza/does-death-exist-new-theo\\_b\\_384515.html](http://www.huffingtonpost.com/robert-lanza/does-death-exist-new-theo_b_384515.html)). *The Huffington Post*. . Retrieved August 30, 2012.
- [11] "Biocentrism Demystified: A Response to Deepak Chopra and Robert Lanza's Notion of a Conscious Universe" (<http://nirmukta.com/2009/12/14/biocentrism-demystified-a-response-to-deepak-chopra-and-robert-lanzas-notion-of-a-conscious-universe/>). nirmukta.com. . Retrieved 2011-01-22.
- [12] Lanza, Robert and Berman, Bob (2009). *Biocentrism: How Life and Consciousness are the Keys to Understanding the True Nature of the Universe* (<http://www.stemcellweblog.com/2009/03/dr-robert-lanzas-book-on-biocentrism.html>). BenBella. ISBN 978-1-933771-69-4. .
- [13] "The Biocentric Universe Theory" ([http://discovermagazine.com/2009/may/01-the-biocentric-universe-life-creates-time-space-cosmos/article\\_view?b\\_start:int=1&-C=/](http://discovermagazine.com/2009/may/01-the-biocentric-universe-life-creates-time-space-cosmos/article_view?b_start:int=1&-C=/)), May 2009 "Discover magazine"
- [14] Lanza and Berman. p. 2.
- [15] Lanza and Berman. p. 17.
- [16] Lanza and Berman. p. 93.
- [17] "A Biotech Provocateur Takes On Physics" ([http://www.forbes.com/2007/03/09/lanza-theories-physics-biotech-oped-cx\\_mh\\_0309lanza.html?partner=yahootix](http://www.forbes.com/2007/03/09/lanza-theories-physics-biotech-oped-cx_mh_0309lanza.html?partner=yahootix)), *Forbes.com*, 9 March 2007
- [18] "Nice debunking of biocentrism" (<http://scienceblogs.com/pharyngula/2009/12/15/nice-debunking-of-biocentrism/>). scienceblogs.com/pharyngula. . Retrieved 2011-01-22.
- [19] "Exclusive: Response to Robert Lanza's essay" ([http://www.usatoday.com/tech/science/2007-03-09-lanza-response\\_N.htm](http://www.usatoday.com/tech/science/2007-03-09-lanza-response_N.htm)). Usatoday.Com. 2007-03-09. . Retrieved 2009-08-17.
- [20] "lanza" (<http://henry.pha.jhu.edu/biocentrism.pdf>) (PDF). . Retrieved 2009-08-17.
- [21] "CNN LIVE EVENT/SPECIAL: CNN International Simulcast: Terrorist Attacks in India (Rush transcript)" (<http://transcripts.cnn.com/TRANSCRIPTS/0811/27/se.04.html>). cnn.com. 2008-11-27. . Retrieved 2009-08-22. ""Our Jonathan Mann posed that question to the Deepak Chopra, the Indian physician and philosopher... DEEPAK CHOPRA, CHOPRA CENTER FOR WELLBEING""
- [22] "Dr. Robert Lanza is Featured Guest on Deepak Chopra's SIRIUS XM Stars Radio Show" (<http://ca.news.finance.yahoo.com/s/13082009/31/link-f-prnewswire-dr-robert-lanza-featured-guest-deepak-chopra-s.html/>). Yahoo Finance. . Retrieved 2009-12-15.

## External links

- Robert Lanza – Biocentrism (<http://www.robertlanzabiocentrism.com/>) Articles, literature, and material on biocentrism
- Biocentricity.net (<http://www.biocentricity.net/>), Videos, FAQ, and news pertaining to biocentrism

# Fine-tuned Universe

---

The **fine-tuned Universe** is the proposition that the conditions that allow life in the Universe can only occur when certain universal fundamental physical constants lie within a very narrow range, so that if any of several fundamental constants were only slightly different, the Universe would be unlikely to be conducive to the establishment and development of matter, astronomical structures, elemental diversity, or life as it is presently understood.<sup>[1]</sup> The existence and extent of fine-tuning in the Universe is a matter of dispute in the scientific community. The proposition is also discussed among philosophers, theologians, creationists, and intelligent design proponents.

Physicist Paul Davies has asserted that "There is now broad agreement among physicists and cosmologists that the Universe is in several respects 'fine-tuned' for life". However, he continues, "the conclusion is not so much that the Universe is fine-tuned for life; rather it is fine-tuned for the building blocks and environments that life requires." He also states that "'anthropic' reasoning fails to distinguish between minimally biophilic universes, in which life is permitted, but only marginally possible, and optimally biophilic universes, in which life flourishes because biogenesis occurs frequently".<sup>[2]</sup> Among scientists who find the evidence persuasive, a variety of natural explanations have been proposed, such as the anthropic principle along with multiple universes.

## History

In 1913, the chemist Lawrence Joseph Henderson (1878–1942) wrote *The Fitness of the Environment*, one of the first books to explore concepts of fine tuning in the Universe. Henderson discusses the importance of water and the environment with respect to living things, pointing out that life depends entirely on the very specific environmental conditions on Earth, especially with regard to the prevalence and properties of water.<sup>[3]</sup>

In 1961, the physicist Robert H. Dicke claimed that certain forces in physics, such as gravity and electromagnetism, must be perfectly fine-tuned for life to exist anywhere in the Universe.<sup>[4][5]</sup> Fred Hoyle also argued for a fine-tuned Universe in his 1984 book *Intelligent Universe*. He compares "the chance of obtaining even a single functioning protein by chance combination of amino acids to a star system full of blind men solving Rubik's Cube simultaneously".<sup>[6]</sup>

John Gribbin and Martin Rees wrote a detailed history and defence of the fine-tuning argument in their book *Cosmic Coincidences* (1989). According to Gribbin and Rees, carbon-based life was not haphazardly arrived at, but the deliberate end of a Universe "tailor-made for man."<sup>[7]</sup>

## Premise

The premise of the fine-tuned Universe assertion is that a small change in several of the dimensionless fundamental physical constants would make the Universe radically different. As Stephen Hawking has noted, "The laws of science, as we know them at present, contain many fundamental numbers, like the size of the electric charge of the electron and the ratio of the masses of the proton and the electron. ... The remarkable fact is that



Fine-tuned Universe proponents argue that deep-space structures such as the Carina Nebula would not form in a universe with minimally different physical constants.

the values of these numbers seem to have been very finely adjusted to make possible the development of life."<sup>[8]</sup>

If, for example, the strong nuclear force were 2% stronger than it is (i.e., if the coupling constant representing its strength were 2% larger), while the other constants were left unchanged, diprotons would be stable and hydrogen would fuse into them instead of deuterium and helium.<sup>[9]</sup> This would drastically alter the physics of stars, and presumably preclude the existence of life similar to what we observe on Earth. The existence of the di-proton would short-circuit the slow fusion of hydrogen into deuterium. Hydrogen would fuse so easily that it is likely that all of the Universe's hydrogen would be consumed in the first few minutes after the Big Bang.<sup>[9]</sup> However, some of the fundamental constants describe the properties of the unstable strange, charmed, bottom and top quarks and mu and tau leptons that seem to play little part in the Universe or the structure of matter.

The precise formulation of the idea is made difficult by the fact that physicists do not yet know how many independent physical constants there are. The current standard model of particle physics has 25 freely adjustable parameters with an additional parameter, the cosmological constant, for gravitation. However, because the standard model is not mathematically self-consistent under certain conditions (e.g., at very high energies, at which both quantum mechanics and general relativity are relevant), physicists believe that it is underlaid by some other theory, such as a grand unified theory, string theory, or loop quantum gravity. In some candidate theories, the actual number of independent physical constants may be as small as one. For example, the cosmological constant may be a fundamental constant, but attempts have also been made to calculate it from other constants, and according to the author of one such calculation, "the small value of the cosmological constant is telling us that a remarkably precise and totally unexpected relation exists among all the parameters of the Standard Model of particle physics, the bare cosmological constant and unknown physics."<sup>[10]</sup>

Martin Rees<sup>[11]</sup> formulates the fine-tuning of the Universe in terms of the following six dimensionless constants:

- $N$  = ratio of the strengths of gravity to that of electromagnetism;
- $Epsilon$  ( $\epsilon$ ) = strength of the force binding nucleons into nuclei;
- $Omega$  ( $\omega$ ) = relative importance of gravity and expansion energy in the Universe;
- $Lambda$  ( $\lambda$ ) = cosmological constant;
- $Q$  = ratio of the gravitational energy required to pull a large galaxy apart to the energy equivalent of its mass;
- $D$  = number of spatial dimensions in spacetime.

## Disputes regarding the existence and extent of fine-tuning

Computer simulations suggest that not all of the purportedly "fine-tuned" parameters may be as fine-tuned as has been claimed. Victor Stenger has simulated different universes in which four fundamental parameters are varied. He found that long-lived stars could exist over a wide parameter range, and concluded that "... a wide variation of constants of physics leads to universes that are long-lived enough for life to evolve, although human life need not exist in such universes".<sup>[12]</sup>

Fred Adams has done a similar study to Stenger, investigating the structure of stars in universes with different values of the gravitational constant  $G$ , the fine-structure constant  $\alpha$ , and a nuclear reaction rate parameter  $C$ . His study suggests that roughly 25% of this parameter space allows stars to exist.<sup>[13]</sup>

The validity of fine tuning examples is sometimes questioned on the grounds that such reasoning is subjective anthropomorphism applied to natural physical constants. Critics also suggest that the fine-tuned Universe assertion and the anthropic principle are essentially tautologies.<sup>[14]</sup> The fine-tuned Universe argument has also been criticized as an argument by lack of imagination, as it assumes no other forms of life, sometimes referred to as carbon chauvinism. Conceptually, alternative biochemistry or other forms of life are possible.<sup>[15]</sup> In addition, critics argue that humans are adapted to the Universe through the process of evolution, rather than the Universe being adapted to humans (see puddle thinking). They also see it as an example of the logical flaw of hubris or anthropocentrism in its assertion that humans are the purpose of the Universe.<sup>[16]</sup>

## Possible naturalistic explanations

There are fine tuning arguments that are naturalistic.<sup>[17]</sup> As modern cosmology developed, various hypotheses have been proposed. One is an oscillatory universe or a multiverse, where fundamental physical constants are postulated to resolve themselves to random values in different iterations of reality.<sup>[18]</sup> Under this hypothesis, separate parts of reality would have wildly different characteristics. In such scenarios, the issue of fine-tuning does not arise at all, as only those "universes" with constants hospitable to life (such as what we observe) would develop life capable of contemplating the question of the origin of fine-tuning.

Based upon the Anthropic principle, physicist Robert H. Dicke proposed the "Dicke coincidence" argument that the structure (age, physical constants, etc.) of the Universe as seen by living observers is not random, but is constrained by biological factors that require it to be roughly a "golden age".<sup>[19]</sup>

## Inflationary cosmology

Inflation theory posits that an inflaton field in the first  $10^{-30}$  seconds of the universe produces strong repulsive gravity, and the universe and space-time expand by a factor of  $10^{30}$ . After  $10^{-30}$  seconds, gravity starts to become attractive. In this framework, with such rapid expansion, the overall shape of the universe at 14 billion years is much less sensitive to initial parameters than the standard big bang model, and thus the fine-tuning issue disappears.<sup>[20]</sup>

## Multiverse

The Multiverse hypothesis assumes the existence of many universes with different physical constants, some of which are hospitable to intelligent life (see multiverse: anthropic principle). Because we are intelligent beings, we are by definition in a hospitable one. Mathematician Michael Ikeda and astronomer William H. Jefferys have argued that the anthropic principle and selection effect resolves the entire issue of fine-tuning,<sup>[21][22]</sup> as does philosopher of science Elliott Sober.<sup>[23]</sup> Philosopher and theologian Richard Swinburne reaches the opposite conclusion using Bayesian probability.<sup>[24]</sup>

This approach has led to considerable research into the anthropic principle and has been of particular interest to particle physicists, because theories of everything do apparently generate large numbers of universes in which the physical constants vary widely. As yet, there is no evidence for the existence of a multiverse, but some versions of the theory do make predictions that some researchers studying M-theory and gravity leaks hope to see some evidence of soon.<sup>[25]</sup> Some multiverse theories are not falsifiable, thus scientists may be reluctant to call any multiverse theory "scientific". UNC-Chapel Hill professor Laura Mersini-Houghton claims that the WMAP cold spot may provide testable empirical evidence for a parallel universe, although this claim was recently refuted as the WMAP cold spot was found to be nothing more than a statistical artifact<sup>[26][27]</sup>

Variants on this approach include Lee Smolin's notion of cosmological natural selection, the Ekpyrotic universe, and the Bubble universe theory.

Critics of the multiverse-related explanations argue that there is no evidence that other universes exist.

## Bubble universe theory

The bubble universe model by physicist Andrei Linde postulates that our Universe is one of many that grew from a multiverse consisting of vacuum that had not yet decayed to its ground state.

According to this scenario, by means of a random quantum fluctuation, the Universe "tunneled" from pure vacuum ("nothing") to what is called a false vacuum, a region of space that contains no matter or radiation, but is not quite "nothing." The space inside this bubble of false vacuum was curved, or warped. A small amount of energy was contained in that curvature, somewhat like the energy stored in a strung bow. This ostensible violation of energy conservation is allowed by the Heisenberg uncertainty principle for sufficiently small time intervals.

The bubble then inflated exponentially and the Universe grew by many orders of magnitude in a tiny fraction of a second. (For a not-too-technical discussion, see Stenger 1990<sup>[28]</sup>). As the bubble expanded, its curvature energy was converted into matter and radiation, inflation stopped, and the more linear Big Bang expansion we now experience commenced. The Universe cooled and its structure spontaneously froze out, as formless water vapor freezes into snowflakes whose unique patterns arise from a combination of symmetry and randomness.

—Victor J. Stenger, *The Anthropic Coincidences*<sup>[29]</sup>

In standard inflation, inflationary expansion occurred while the Universe was in a false vacuum state, halting when the Universe decayed to a true vacuum state. The *bubble universe* model proposes that different parts of this inflationary universe (termed a Multiverse) decayed at different times, with decaying regions corresponding to universes not in causal contact with each other. It further supposes that each bubble universe may have different physical constants.

### Top-down cosmology

Stephen Hawking, along with Thomas Hertog of CERN, proposed that the Universe's initial conditions consisted of a superposition of many possible initial conditions, only a small fraction of which contributed to the conditions we see today.<sup>[30]</sup> According to their theory, it is inevitable that we find our Universe's "fine-tuned" physical constants, as the current Universe "selects" only those past histories that led to the present conditions. In this way, top-down cosmology provides an anthropic explanation for why we find ourselves in a universe that allows matter and life, without invoking the existence of the Multiverse.<sup>[31]</sup>

### Alien design

One hypothesis is that the Universe may have been designed by extra-universal aliens. Some believe this would solve the problem of how a designer or design team capable of fine-tuning the Universe could come to exist. Cosmologist Alan Guth believes humans will in time be able to generate new universes. By implication previous intelligent entities may have generated our Universe.<sup>[32]</sup> This idea leads to the possibility that the extraterrestrial designer/designers are themselves the product of an evolutionary process in their own universe, which must therefore itself be able to sustain life. However it also raises the question of where this universe came from, leading to an infinite regress.

The Simulation hypothesis promoted by Nick Bostrom and others suggests that our Universe may be a computer simulation by aliens.<sup>[33]</sup>

The Biocosm hypothesis and the Meduso-anthropic principle both suggest that natural selection has made the universe biophilic. The Universe enables intelligence because intelligent entities later create new biophilic universes. This is different from the suggestion above that aliens from a universe that is less-finely tuned than ours made our Universe finely tuned.

The Designer Universe theory of John Gribbin suggests that the Universe could have been made deliberately by a member or members of a technologically advanced civilization in another part of the Multiverse, and that this advanced civilization may have been responsible for causing the Big Bang.<sup>[34]</sup>

## Religious arguments

As with theistic evolution, some individual scientists, theologians, and philosophers as well as certain religious groups argue that providence or creation are responsible for fine-tuning.

Christian philosopher Alvin Plantinga argues that random chance, applied to a single and sole universe, only raises the question as to why this universe could be so "lucky" as to have precise conditions that support life at least at some place (the Earth) and time (within millions of years of the present).

One reaction to these apparent enormous coincidences is to see them as substantiating the theistic claim that the Universe has been created by a personal God and as offering the material for a properly restrained theistic argument—hence the fine-tuning argument. It's as if there are a large number of dials that have to be tuned to within extremely narrow limits for life to be possible in our Universe. It is extremely unlikely that this should happen by chance, but much more likely that this should happen, if there is such a person as God.

—Alvin Plantinga, *The Dawkins Confusion; Naturalism ad absurdum*<sup>[35]</sup>

This fine-tuning of the Universe is cited<sup>[36]</sup> by theologian and philosopher William Lane Craig as an evidence for the existence of God or some form of intelligence capable of manipulating (or designing) the basic physics that governs the Universe. Craig argues, however, "that the postulate of a divine Designer does not settle for us the religious question."

## Intelligent design

Proponents of Intelligent design argue that certain features of the Universe and of living things are best explained by an intelligent cause, not an undirected process such as natural selection. The fine-tuned Universe argument is a central premise or presented as given in many of the published works of prominent intelligent design proponents, such as William A. Dembski and Michael Behe.

## Other religious creation views

Most religions have some kind of account of the creation of the Universe, although they generally differ in detail from the ones listed above. Some of these may be compatible with known scientific facts (Old Earth Creationism, Theistic Evolution, Progressive Creationism). Others are either incompatible with, or indifferent to, scientific understandings (Young Earth Creationism). For example scientist-theologians such as John Polkinghorne emphasize the implications of *Anthropic Fine-Tuning* within an orthodox Christian framework, whilst fully accepting the scientific findings about Evolution and the age of the Universe. This is also the position of the Roman Catholic Church and of most Anglican theologians.<sup>[37]</sup> The Jewish physicist Gerald Schroeder argues that the apparent discrepancy between the "days" in Genesis and the billions of years in a scientific understanding are due to the differences in frames of reference. Other scientists with similar views are physicist Freeman Dyson and astronomer Owen Gingerich.

## Counter argument to religious views

Victor Stenger argues that "... The fine-tuning argument and other recent intelligent design arguments are modern versions of God-of-the-gaps reasoning, where a God is deemed necessary whenever science has not fully explained some phenomenon".<sup>[12]</sup>

The argument from imperfection suggests that if the Universe were designed to be fine-tuned for life, it should be the best one possible and that evidence suggests that it is not.<sup>[38]</sup> In fact, most of the Universe is highly hostile to life.

Additionally, Stenger argues: "We have no reason to believe that our kind of carbon-based life is all that is possible. Furthermore, modern cosmology indicates that multiple universes may exist with different constants and laws of physics. So, it is not surprising that we live in the one suited for us. The Universe is not fine-tuned to life; life is fine-tuned to the Universe."<sup>[39]</sup>

## In popular culture

- John C. Lennox discusses the fine-tuned Universe at length in his novel *GOD'S UNDERTAKER: HAS SCIENCE BURIED GOD?* (2006).
- Robert J. Sawyer discusses the fine-tuned Universe at length in his novel *Calculating God* (2000).
- Author Neal Stephenson discussed the issue of fine-tuning in the conclusion to his essay *In the Beginning... was the Command Line*.<sup>[40]</sup>
- **Puddle thinking** is a satirical illustration of the "*life is fine-tuned to the Universe*" argument above coined by Douglas Adams to satirize the Fine-tuned Universe argument for supernatural creationism.<sup>[41][42]</sup> As quoted in Richard Dawkins' eulogy for Douglas Adams:<sup>[43]</sup>

... imagine a puddle waking up one morning and thinking, 'This is an interesting world I find myself in, an interesting hole I find myself in, fits me rather neatly, doesn't it? In fact, it fits me staggeringly well, must have been made to have me in it!' This is such a powerful idea that as the Sun rises in the sky and the air heats up and as, gradually, the puddle gets smaller and smaller, it's still frantically hanging on to the notion that everything's going to be all right, because this World was meant to have him in it, was built to have him in it; so the moment he disappears catches him rather by surprise. I think this may be something we need to be on the watch out for.

## References

- [1] Mark Isaak (ed.) (2005). "CI301: The Anthropic Principle" (<http://www.talkorigins.org/indexcc/CI/CI301.html>). *Index to Creationist Claims*. TalkOrigins Archive. . Retrieved 2007-10-31.
- [2] Davies (2003). "How bio-friendly is the universe". *Int.J.Astrobiol* **2** (115): 115. arXiv:astro-ph/0403050. doi:10.1017/S1473550403001514.
- [3] Lawrence Joseph Henderson, *The fitness of the environment: an inquiry into the biological significance of the properties of matter* The Macmillan Company, 1913
- [4] Robert H. Dicke, "Dirac's Cosmology and Mach's Principle," *Nature* 192 (1961): 440-41
- [5] Heilbron, J. L. *The Oxford guide to the history of physics and astronomy, Volume 10* 2005, p. 8
- [6] Profile of Fred Hoyle at OPT (<http://www.optcorp.com/edu/articleDetailEDU.aspx?aid=1530>)
- [7] Gribbin. J and Rees. M, *Cosmic Coincidences: Dark Matter, Mankind, and Anthropic Cosmology*, 1989, ISBN 0-553-34740-3
- [8] Stephen Hawking, 1988. *A Brief History of Time*, Bantam Books, ISBN 0-553-05340-X, p. 125.
- [9] Paul Davies, 1993. *The Accidental Universe*, Cambridge University Press, p70-71
- [10] Larry Abbott, "The Mystery of the Cosmological Constant," *Scientific American*, vol. 3, no. 1 (1991): 78.
- [11] Martin Rees, 1999. *Just Six Numbers*, HarperCollins Publishers, ISBN 0-465-03672-4.
- [12] Is The Universe Fine-Tuned For Us? ([http://www.colorado.edu/philosophy/vstenger/Cosmo/FineTune.pdf#search="Fine tuned universe"](http://www.colorado.edu/philosophy/vstenger/Cosmo/FineTune.pdf#search=)) Victor J. Stenger, University of Colorado.
- [13] Adams, F.C. (2008). "Stars in other universes: stellar structure with different fundamental constants". *Journal of Cosmology and Astroparticle Physics* **2008** (8): 010. arXiv:0807.3697. Bibcode 2008JCAP...08..010A. doi:10.1088/1475-7516/2008/08/010.
- [14] See, e.g., Our place in the Multiverse (<http://www.nature.com/nature/journal/v443/n7108/full/443145a.html>) Joseph Silk. *Nature*, Volume 443 Number 7108, September 14, 2006.
- [15] See, e.g. Jack Cohen and Ian Stewart: *What Does a Martian Look Like: The Science of Extraterrestrial Life*, Wiley, 2002
- [16] See, e.g., Gerald Feinberg and Robert Shapiro, "A Puddlian Fable" in Huchingson, *Religion and the Natural Sciences* (1993), pp. 220-221
- [17] L. Susskind, *The cosmic landscape: string theory and the illusion of intelligent design*. (Little, Brown, 2005)
- [18] Wheeler, J. A. (1977) in *Foundational problems in the special sciences*, Reidel, Dordrecht, pp 3–33
- [19] Dicke, R. H. (1961). "Dirac's Cosmology and Mach's Principle". *Nature* **192** (4801): 440–441. Bibcode 1961Natur.192..440D. doi:10.1038/192440a0.
- [20] The Fabric of the Cosmos ([http://www.amazon.co.uk/The-Fabric-Cosmos-Texture-Reality/dp/0141011114/ref=sr\\_1\\_1?ie=UTF8&qid=1338294220&sr=8-1](http://www.amazon.co.uk/The-Fabric-Cosmos-Texture-Reality/dp/0141011114/ref=sr_1_1?ie=UTF8&qid=1338294220&sr=8-1)).
- [21] The Anthropic Principle Does Not Support Supernaturalism (<http://quasar.as.utexas.edu/anthropic.html>), Michael Ikeda, Bill Jefferys
- [22] Michael Ikeda and William H. Jefferys, "The Anthropic Principle Does Not Support Supernaturalism," in *The Improbability of God*, Michael Martin and Ricki Monnier, Editors, pp. 150-166. Amherst, N.Y.: Prometheus Press. ISBN 1-59102-381-5.
- [23] Elliott Sober, 2004. *The Design Argument*, in *The Blackwell Guide to the Philosophy of Religion*, W. E. Mann, Editor. Blackwell Publishing, ISBN 0-631-22129-8.
- [24] Richard Swinburne, 1990. *Argument from the fine-tuning of the Universe*, in *Physical cosmology and philosophy*, J. Leslie, Editor. Collier Macmillan: New York. pp. 154-73.
- [25] *Parallel Worlds*, 2005, Michio Kaku, pp. 220-221



- [26] <http://www.universetoday.com/40413/what-no-parallel-universe-cosmic-cold-spot-just-data-artifact/>
- [27] "Two Programmes - Horizon, 2010-2011, What Happened Before the Big Bang?" (<http://www.bbc.co.uk/programmes/b00vdkmj>). BBC. . Retrieved 2011-01-02.
- [28] Stenger, Victor J. 1990. "The Universe: The Ultimate Free Lunch." *European Journal of Physics* 11, 236. (<http://iopscience.iop.org/0143-0807/11/4/008;jsessionid=EA40A0D8683BE1342C049CC44EDAFB4E.c3>)
- [29] The Anthropic Coincidences ([http://www.colorado.edu/philosophy/vstenger/Cosmo/anthro\\_skintel.html](http://www.colorado.edu/philosophy/vstenger/Cosmo/anthro_skintel.html))
- [30] Ball, Philip (June 21, 2006). "Hawking Rewrites History...Backwards" (<http://www.nature.com/news/2006/060619/full/news060619-6.html>). *Nature News Online*. . Retrieved April 19, 2010.
- [31] Hawking, S. W.; Hertog, Thomas (February 2006). "Populating the Landscape: A Top Down Approach" (<http://arxiv.org/abs/hep-th/0602091>). *Phys. Rev. D* **73** (12): 123527. arXiv:hep-th/0602091v2. Bibcode 2006PhRvD..7313527H. doi:10.1103/PhysRevD.73.123527. . Retrieved June 15, 2012.
- [32] BBC - Science & Nature - Horizon - Parallel Universes - Transcript (<http://www.bbc.co.uk/science/horizon/2001/parallelunitrans.shtml>)
- [33] Bostrom, N. (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge, New York. ISBN 0-415-93858-9.
- [34] John Gribbin, *In Search of the Multiverse: Parallel Worlds, Hidden Dimensions, and the Ultimate Quest for the Frontiers of Reality*, 2010, p. 195
- [35] Alvin Plantinga, "The Dawkins Confusion; Naturalism ad absurdum," *Christianity Today*, March/April 2007 (<http://www.philvaz.com/apologetics/DawkinsGodDelusionPlantingaReview.pdf>)
- [36] by William Lane Craig, "The Teleological Argument and the Anthropic Principle," (<http://www.leaderu.com/offices/billcraig/docs/teleo.html>)
- [37] See, e.g., Alister McGrath's books *Scientific Theology* and *The Science of God*.
- [38] Avitel Pilpel, SKEPTIC, November 2007 Issue, p.18
- [39] Victor Stenger, Flew's Flawed Science (<http://www.mukto-mona.com/Articles/vstenger/flew.htm>)
- [40] In The Beginning Was The Command Line (<http://www.cryptonomicon.com/beginning.html>)
- [41] Williams, Robyn (18). "The anthropic universe" (<http://www.abc.net.au/rn/scienceshow/stories/2006/1572643.htm>). *The Science Show* (ABC Radio National). . Retrieved 19 November 2009.
- [42] Redfern, Martin (24 December 1995). "Proofs of God in a photon" (<http://www.independent.co.uk/arts-entertainment/proofs-of-god-in-a-photon-1527306.html>). *The Independent*. . Retrieved 19 November 2009.
- [43] Dawkins, Richard (17 September 2001). "Eulogy for Douglas Adams" ([http://www.edge.org/documents/adams\\_index.html](http://www.edge.org/documents/adams_index.html)). *Edge*. . Retrieved 19 November 2009.

## Further reading

- <sup>[</sup> Edit this reference (<http://en.wikipedia.org/w/index.php?title=Template:BarrowTipler1986&action=edit>)<sup>]</sup>

Barrow, John D.; Tipler, Frank J. (19 May 1988). *The Anthropic Cosmological Principle* (<http://books.google.com/books?id=uSykSbXklWEC&printsec=frontcover>). foreword by John A. Wheeler. Oxford: Oxford University Press. ISBN 9780192821478. LC 87-28148 (<http://lccn.loc.gov/87028148>). Retrieved 31 December 2009.

- John D. Barrow, 2003. *The Constants of Nature*, Pantheon Books, ISBN 0-375-42221-8
- Bernard Carr, ed. (2007) *Universe or Multiverse?* Cambridge University Press.
- Paul Davies, 1982. *The Accidental Universe*, Cambridge University Press, ISBN 0-521-24212-6
- Paul Davies, 2007. *Cosmic Jackpot: Why Our Universe Is Just Right for Life*, Houghton Mifflin Harcourt, ISBN 0-618-59226-1. Reprinted as: *The Goldilocks Enigma: Why Is the Universe Just Right for Life?*, 2008, Mariner Books, ISBN 0-547-05358-4.
- Alister McGrath, 2009. *A Fine-Tuned Universe: The Quest for God in Science and Theology*, Westminster John Knox Press, ISBN 0-664-23310-4.
- Simon Conway Morris, 2003. *Life's Solution: Inevitable Humans in a Lonely Universe*. Cambridge Univ. Press.
- Peter Ward and Donald Brownlee, 2000. *Rare Earth: Why Complex Life is Uncommon in the Universe*. Springer Verlag.

## External links

### Defend fine-tuning:

- God: new evidence – a series of six videos exploring how evidence from cosmic fine tuning points to the reality of a creator God (<http://www.focus.org.uk/>)
- Stephen M. Barr (2001) " Overview of the Cosmological Intelligent Design argument, (<http://www.firstthings.com/ftissues/ft0106/articles/barr.html>)" *First Things, the Journal of Religion, Culture, and Public Life*.
- Robin Collins:
  - Fine-Tuning website. (<http://home.messiah.edu/~rcollins/>)
  - " Cosmological fine-tuning. ([http://academic.udayton.edu/WilliamRichards/Intro essays/Collins, Fine-tuning.htm](http://academic.udayton.edu/WilliamRichards/Intro%20essays/Collins,%20Fine-tuning.htm))"
- Hugh Ross:
  - " Design and the Anthropic Principle. (<http://www.reasons.org/resources/apologetics/design.shtml?main>)"
  - " Evidence For Design In The Universe. (<http://doesgodexist.com/Charts/EvidenceForDesignInTheUniverse.html>)"
- Gerald Schroeder: Fine-tuned universe. (<http://www.geraldschroeder.com/FineTuning.aspx>)
- The cosmos is fine-tuned to permit human life (<http://www.talkorigins.org/indexcc/CI/CI301.html>) at the talk.origins index to creationist claims.
- Interview ([http://www.berkeley.edu/news/media/releases/2005/06/17\\_townes.shtml](http://www.berkeley.edu/news/media/releases/2005/06/17_townes.shtml)) with Charles Townes discussing science and religion.
- Home page (<http://www.templeton.org/biochem-finetuning/purpose.html>) of Templeton Foundation project on fine-tuning.

### Criticize fine tuning:

- Bibliography of online Links to criticisms of the Fine-Tuning Argument. (<http://www.infidels.org/library/modern/theism/design.html#fine>) Secular Web.
- Ikeda, Michael, and Jefferys, William H., " The Anthropic Principle Does Not Support Supernaturalism. (<http://quasar.as.utexas.edu/anthropic.html>)"
- Jefferys, W. H., and J. O. Berger, " Sharpening Ockham's razor on a Bayesian stop. (<http://bayesrules.net/papers/ockham.pdf>)"
- Victor Stenger:
  - " Does the Cosmos Show Evidence of Purpose? (<http://www.colorado.edu/philosophy/vstenger/Cosmo/anthro.skinq.html>)"
  - " Is the Universe fine-tuned for us? (<http://www.colorado.edu/philosophy/vstenger/Cosmo/FineTune.pdf>)"
- Elliott Sober, " The Design Argument. ([http://philosophy.wisc.edu/sober/design argument 11 2004.pdf](http://philosophy.wisc.edu/sober/design%20argument%2011%202004.pdf))" An earlier version appeared in the *Blackwell Companion to the Philosophy of Religion* (2004).

# Conservation of energy

The **law of conservation of energy**, first formulated in the nineteenth century, is a law of physics. It states that the total amount of energy in an isolated system remains constant over time. The total energy is said to be *conserved* over time. For an isolated system, this law means that energy can change its location within the system, and that it can change form within the system, for instance chemical energy can become kinetic energy, but that energy can be neither created nor destroyed.

In the twentieth century, the definition of energy was broadened. It was found that particles that have rest mass are equivalent to amounts of energy (see mass-energy equivalence). These particles were found subject to annihilation in which *matter* particles (such as electrons) can be converted to *non-matter* (such as photons of electromagnetic radiation), or even into potential energy or kinetic energy. Matter could also be created out of kinetic or other types of energy, in the process of matter creation. Thus, matter (defined as ponderable matter particles) was found not to be conserved.

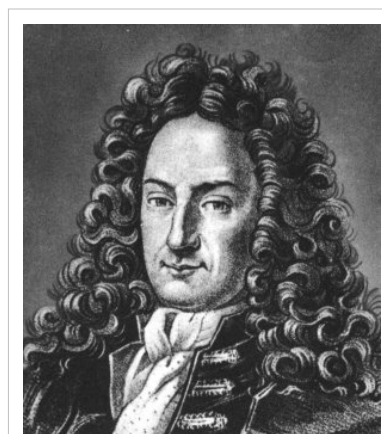
In such a transformation process within an isolated system, neither the *mass* nor the *energy* changes over time, although the matter content may change. Therefore, conservation of energy, and conservation of mass, each still holds as a law in its own right (indeed they are restatements of the same law, when mass and energy are recognized to be equivalent). When stated alternatively, in terms of mass and of energy, they appear as the apparently distinct laws of the nineteenth century.

A consequence of the law of conservation of energy is that no intended "perpetual motion machine" can perpetually deliver energy to its surroundings.<sup>[1]</sup> Any delivery of energy by such a device would result in delivery of mass also, and the machine would lose mass continually until it eventually disappeared.

## History

Ancient philosophers as far back as Thales of Miletus c.~550 BCE had inklings of the conservation of some underlying substance of which everything is made. However, there is no particular reason to identify this with what we know today as "mass-energy" (for example, Thales thought it was water). Empedocles (490–430 BCE) wrote that in his universal system, composed of four roots (earth, air, water, fire), "nothing comes to be or perishes",<sup>[2]</sup> but these elements suffer continual rearrangement.

In 1638, Galileo published his analysis of several situations—including the celebrated "interrupted pendulum"—which can be described (in modern language) as conservatively converting potential energy to kinetic energy and back again. However, Galileo did not state the process in modern terms and again cannot be credited with the crucial insight.



Gottfried Leibniz

It was Gottfried Wilhelm Leibniz during 1676–1689 who first attempted a mathematical formulation of the kind of energy which is connected with *motion* (kinetic energy). Leibniz noticed that in many mechanical systems (of several masses,  $m_i$  each with velocity  $v_i$ ),

$$\sum_i m_i v_i^2$$

was conserved so long as the masses did not interact. He called this quantity the *vis viva* or *living force* of the system. The principle represents an accurate statement of the approximate conservation of kinetic energy in situations where there is no friction. Many physicists at that time held that the conservation of momentum, which holds even in systems with friction, as defined by the momentum:

$$\sum_i m_i v_i$$

was the conserved *vis viva*. It was later shown that, under the proper conditions, both quantities are conserved simultaneously such as in elastic collisions.

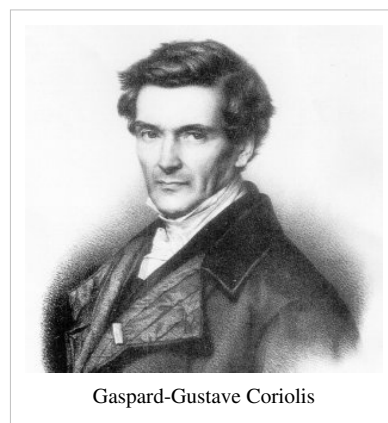
It was largely engineers such as John Smeaton, Peter Ewart, Carl Holtzmann, Gustave-Adolphe Hirn and Marc Seguin who objected that conservation of momentum alone was not adequate for practical calculation and made use of Leibniz's principle. The principle was also championed by some chemists such as William Hyde Wollaston. Academics such as John Playfair were quick to point out that kinetic energy is clearly not conserved. This is obvious to a modern analysis based on the second law of thermodynamics but in the 18th and 19th centuries, the fate of the lost energy was still unknown. Gradually it came to be suspected that the heat inevitably generated by motion under friction, was another form of *vis viva*. In 1783, Antoine Lavoisier and Pierre-Simon Laplace reviewed the two competing theories of *vis viva* and caloric theory.<sup>[3]</sup> Count Rumford's 1798 observations of heat generation during the boring of cannons added more weight to the view that mechanical motion could be converted into heat, and (as importantly) that the conversion was quantitative and could be predicted (allowing for a universal conversion constant between kinetic energy and heat). *Vis viva* now started to be known as *energy*, after the term was first used in that sense by Thomas Young in 1807.

The recalibration of *vis viva* to

$$\frac{1}{2} \sum_i m_i v_i^2$$

which can be understood as finding the exact value for the kinetic energy to work conversion constant, was largely the result of the work of Gaspard-Gustave Coriolis and Jean-Victor Poncelet over the period 1819–1839. The former called the quantity *quantité de travail* (quantity of work) and the latter, *travail mécanique* (mechanical work), and both championed its use in engineering calculation.

In a paper *Über die Natur der Wärme*, published in the *Zeitschrift für Physik* in 1837, Karl Friedrich Mohr gave one of the earliest general statements of the doctrine of the conservation of energy in the words: "besides the 54 known chemical elements there is in the physical world one agent only, and this is called *Kraft* [energy or work]. It may appear, according to circumstances, as motion, chemical affinity, cohesion, electricity, light and magnetism; and from any one of these forms it can be transformed into any of the others."

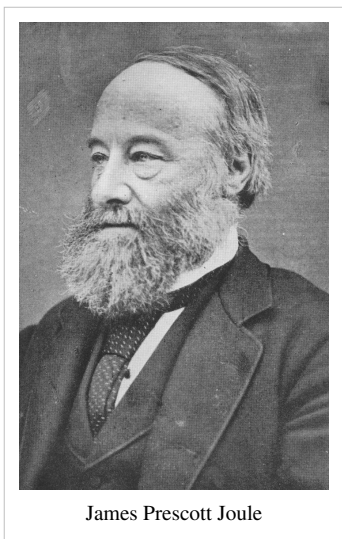


Gaspard-Gustave Coriolis

### Mechanical equivalent of heat

A key stage in the development of the modern conservation principle was the demonstration of the *mechanical equivalent of heat*. The caloric theory maintained that heat could neither be created nor destroyed but conservation of energy entails the contrary principle that heat and mechanical work are interchangeable.

In 1798 Count Rumford (Benjamin Thompson) performed measurements of the frictional heat generated in boring cannons and developed the idea that heat is a form of kinetic energy; his measurements refuted caloric theory, but were imprecise enough to leave room for doubt.



James Prescott Joule

The mechanical equivalence principle was first stated in its modern form by the German surgeon Julius Robert von Mayer in 1842.<sup>[4]</sup> Mayer reached his conclusion on a voyage to the Dutch East Indies, where he found that his patients' blood was a deeper red because they were consuming less oxygen, and therefore less energy, to maintain their body temperature in the hotter climate. He discovered that heat and mechanical work were both forms of energy and in 1845, after improving his knowledge of physics, he published a monograph that stated a quantitative relationship between them.<sup>[5]</sup>

Meanwhile, in 1843 James Prescott Joule independently discovered the mechanical equivalent in a series of experiments. In the most famous, now called the "Joule apparatus", a descending weight attached to a string caused a paddle immersed in water to rotate. He showed that the gravitational potential energy lost by the weight in descending was equal to the thermal energy (heat) gained by the water by friction with the paddle.

Over the period 1840–1843, similar work was carried out by engineer Ludwig A. Colding though it was little known outside his native Denmark.

Both Joule's and Mayer's work suffered from resistance and neglect but it was Joule's that eventually drew the wider recognition.

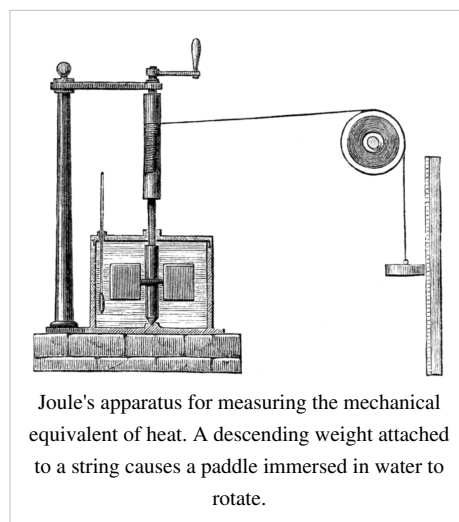
*For the dispute between Joule and Mayer over priority, see Mechanical equivalent of heat: Priority*

In 1844, William Robert Grove postulated a relationship between mechanics, heat, light, electricity and magnetism by treating them all as manifestations of a single "force" (*energy* in modern terms). In 1874 Grove published his theories in his book *The Correlation of Physical Forces*.<sup>[6]</sup> In 1847, drawing on the earlier work of Joule, Sadi Carnot and Émile Clapeyron, Hermann von Helmholtz arrived at conclusions similar to Grove's and published his theories in his book *Über die Erhaltung der Kraft (On the Conservation of Force, 1847)*. The general modern acceptance of the principle stems from this publication.

In 1877, Peter Guthrie Tait claimed that the principle originated with Sir Isaac Newton, based on a creative reading of propositions 40 and 41 of the *Philosophiae Naturalis Principia Mathematica*. This is now regarded as an example of Whig history.<sup>[7]</sup>

## Mass–energy equivalence

In the nineteenth century, mass and energy were considered to be of quite different natures. Then Albert Einstein's theory of special relativity showed that mass and energy are related by an equivalence. Energy has an equivalent mass, and mass has an equivalent energy. Physicists now speak of a unified law of conservation of *mass-energy*. This is a recognition that the two nineteenth century conservation laws are restricted versions of one and the same more general law. While matter can be actually converted into non-matter, the relation between mass and energy is a simply theoretical equivalence, so that it makes no sense to think of their "actual interconversion". Thus, the modern



Joule's apparatus for measuring the mechanical equivalent of heat. A descending weight attached to a string causes a paddle immersed in water to rotate.

view is that conservation of energy and conservation of mass are simply the same conservation law, stated differently, in different units. Einstein's  $E = mc^2$  and other equations serve to convert one unit to the other.

## First law of thermodynamics

For a closed thermodynamic system, the first law of thermodynamics may be stated as:

$$\delta Q = dU + \delta W, \text{ or equivalently, } dU = \delta Q - \delta W,$$

where  $\delta Q$  is the amount of energy added to the system by a heating process,  $\delta W$  is the amount of energy lost by the system due to work done by the system on its surroundings and  $dU$  is the change in the internal energy of the system.

The  $\delta$ 's before the heat and work terms are used to indicate that they describe an increment of energy which is to be interpreted somewhat differently than the  $dU$  increment of internal energy (see Inexact differential). Work and heat are *processes* which add or subtract energy, while the internal energy  $U$  is a particular *form* of energy associated with the system. Thus the term "heat energy" for  $\delta Q$  means "that amount of energy added as the result of heating" rather than referring to a particular form of energy. Likewise, the term "work energy" for  $\delta W$  means "that amount of energy lost as the result of work". The most significant result of this distinction is the fact that one can clearly state the amount of internal energy possessed by a thermodynamic system, but one cannot tell how much energy has flowed into or out of the system as a result of its being heated or cooled, nor as the result of work being performed on or by the system. In simple terms, this means that energy cannot be created or destroyed, only converted from one form to another.

Entropy is a function of the state of a system which tells of the possibility of conversion of heat into work.

For a simple compressible system, the work performed by the system may be written:

$$\delta W = P dV,$$

where  $P$  is the pressure and  $dV$  is a small change in the volume of the system, each of which are system variables.

The heat energy may be written

$$\delta Q = T dS,$$

where  $T$  is the temperature and  $dS$  is a small change in the entropy of the system. Temperature and entropy are variables of state of a system.

For a simple open system (in which the particles may be exchanged with the environment) containing a single type of particle, the first law is written:

$$dU = \delta Q - \delta W + \mu dN,$$

where  $dN$  is the change in the number of particles and  $\mu$  is the chemical potential per particle, the energy per added particle required to maintain an unchanged volume and entropy.

## Mechanics

In mechanics, conservation of energy is usually stated as

$$E = T + V,$$

where  $T$  is kinetic energy and  $V$  potential energy.

For this particular form to be valid, the following must be true:

- The system is scleronomous (neither kinetic nor potential energy are explicit functions of time)
- The potential energy doesn't depend on velocities.
- The kinetic energy is a quadratic form with regard to velocities.
- The total energy  $E$  depends on the motion of the frame of reference (and it turns out that it is minimum for the center of mass frame).

## Noether's theorem

The conservation of energy is a common feature in many physical theories. From a mathematical point of view it is understood as a consequence of Noether's theorem, which states every continuous symmetry of a physical theory has an associated conserved quantity; if the theory's symmetry is time invariance then the conserved quantity is called "energy". The energy conservation law is a consequence of the shift symmetry of time; energy conservation is implied by the empirical fact that the laws of physics do not change with time itself. Philosophically this can be stated as "nothing depends on time per se". In other words, if the physical system is invariant under the continuous symmetry of time translation then its energy (which is canonical conjugate quantity to time) is conserved. Conversely, systems which are not invariant under shifts in time (an example, systems with time dependent potential energy) do not exhibit conservation of energy – unless we consider them to exchange energy with another, external system so that the theory of the enlarged system becomes time invariant again. Since any time-varying system can be embedded within a larger time-invariant system, conservation can always be recovered by a suitable re-definition of what energy is. Conservation of energy for finite systems is valid in such physical theories as special relativity and quantum theory (including QED) in the flat space-time.

## Relativity

With the discovery of special relativity by Albert Einstein, energy was proposed to be one component of an energy-momentum 4-vector. Each of the four components (one of energy and three of momentum) of this vector is separately conserved across time, in any closed system, as seen from any given inertial reference frame. Also conserved is the vector length (Minkowski norm), which is the rest mass for single particles, and the invariant mass for systems of particles (where momenta and energy are separately summed before the length is calculated—see the article on invariant mass).

The relativistic energy of a single massive particle contains a term related to its rest mass in addition to its kinetic energy of motion. In the limit of zero kinetic energy (or equivalently in the rest frame) of a massive particle; or else in the center of momentum frame for objects or systems which retain kinetic energy, the total energy of particle or object (including internal kinetic energy in systems) is related to its rest mass or its invariant mass via the famous equation  $E = mc^2$ .

Thus, the rule of *conservation of energy* over time in special relativity continues to hold, so long as the reference frame of the observer is unchanged. This applies to the total energy of systems, although different observers disagree as to the energy value. Also conserved, and invariant to all observers, is the invariant mass, which is the minimal system mass and energy that can be seen by any observer, and which is defined by the energy–momentum relation.

In general relativity conservation of energy-momentum is expressed with the aid of a stress-energy-momentum pseudotensor. The theory of general relativity leaves open the question of whether there is a conservation of energy for the entire universe.

## Quantum theory

In quantum mechanics, energy of a quantum system is described by a self-adjoint (Hermite) operator called Hamiltonian, which acts on the Hilbert space (or a space of wave functions ) of the system. If the Hamiltonian is a time independent operator, emergence probability of the measurement result does not change in time over the evolution of the system. Thus the expectation value of energy is also time independent. The local energy conservation in quantum field theory is ensured by the quantum Noether's theorem for energy-momentum tensor operator. Note that due to the lack of the (universal) time operator in quantum theory, the uncertainty relations for time and energy are not fundamental in contrast to the position momentum uncertainty principle, and merely holds in specific cases (See Uncertainty principle). Energy at each fixed time can be precisely measured in principle without any problem caused by the time energy uncertainty relations. Thus the conservation of energy in time is a well defined concept even in quantum mechanics.

## Notes

- [1] Planck, M. (1923/1927). *Treatise on Thermodynamics*, third English edition translated by A. Ogg from the seventh German edition, Longmans, Green & Co., London, page 40.
- [2] Janko, Richard (2004). "Empedocles, "On Nature"". *Zeitschrift für Papyrologie und Epigraphik*.
- [3] Lavoisier, A.L. & Laplace, P.S. (1780) "Memoir on Heat", *Académie Royale des Sciences* pp. 4–355
- [4] von Mayer, J.R. (1842) "Remarks on the forces of inorganic nature" in *Annalen der Chemie und Pharmacie*, **43**, 233
- [5] Mayer, J.R. (1845). *Die organische Bewegung in ihrem Zusammenhange mit dem Stoffwechsel. Ein Beitrag zur Naturkunde*, Dechslers, Heilbronn.
- [6] Grove, W. R. (1874). *The Correlation of Physical Forces* (6th ed.). London: Longmans, Green.
- [7] Hadden, Richard W. (1994). *On the shoulders of merchants: exchange and the mathematical conception of nature in early modern Europe* (<http://books.google.com/books?id=7IxtC4Jw1YoC>). SUNY Press. p. 13. ISBN 0-7914-2011-6. , Chapter 1, p. 13 (<http://books.google.com/books?id=7IxtC4Jw1YoC&pg=PA13>)

## References

### Modern accounts

- Goldstein, Martin, and Inge F., 1993. *The Refrigerator and the Universe*. Harvard Univ. Press. A gentle introduction.
- Kroemer, Herbert; Kittel, Charles (1980). *Thermal Physics (2nd ed.)*. W. H. Freeman Company. ISBN 0-7167-1088-9.
- Nolan, Peter J. (1996). *Fundamentals of College Physics, 2nd ed.*. William C. Brown Publishers.
- Oxtoby & Nachtrieb (1996). *Principles of Modern Chemistry*, 3rd ed.. Saunders College Publishing.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
- Serway, Raymond A.; Jewett, John W. (2004). *Physics for Scientists and Engineers (6th ed.)*. Brooks/Cole. ISBN 0-534-40842-7.
- Stenger, Victor J. (2000). *Timeless Reality*. Prometheus Books. Especially chpt. 12. Nontechnical.
- Tipler, Paul (2004). *Physics for Scientists and Engineers: Mechanics, Oscillations and Waves, Thermodynamics (5th ed.)*. W. H. Freeman. ISBN 0-7167-0809-4.
- Lanczos, Cornelius (1970). *The Variational Principles of Mechanics*. Toronto: University of Toronto Press. ISBN 0-8020-1743-6.

### History of ideas

- Brown, T.M. (1965). "Resource letter EEC-1 on the evolution of energy concepts from Galileo to Helmholtz". *American Journal of Physics* **33** (10): 759–765. Bibcode 1965AmJPh..33..759B. doi:10.1119/1.1970980.
- Cardwell, D.S.L. (1971). *From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age*. London: Heinemann. ISBN 0-435-54150-1.
- Guillen, M. (1999). *Five Equations That Changed the World*. New York: Abacus. ISBN 0-349-11064-6.
- Hiebert, E.N. (1981). *Historical Roots of the Principle of Conservation of Energy*. Madison, Wis.: Ayer Co Pub. ISBN 0-405-13880-6.
- Kuhn, T.S. (1957) "Energy conservation as an example of simultaneous discovery", in M. Clagett (ed.) *Critical Problems in the History of Science* pp.321–56
- Sarton, G.; Joule, J. P.; Carnot, Sadi (1929). "The discovery of the law of conservation of energy". *Isis* **13**: 18–49. doi:10.1086/346430.
- Smith, C. (1998). *The Science of Energy: Cultural History of Energy Physics in Victorian Britain*. London: Heinemann. ISBN 0-485-11431-3.
- Mach, E. (1872). *History and Root of the Principles of the Conservation of Energy*. Open Court Pub. Co., IL.
- Poincaré, H. (1905). *Science and Hypothesis*. Walter Scott Publishing Co. Ltd; Dover reprint, 1952. ISBN 0-486-60221-4., Chapter 8, "Energy and Thermo-dynamics"



## External links

- MISN-0-158 *The First Law of Thermodynamics* ([http://35.9.69.219/home/modules/pdf\\_modules/m158.pdf](http://35.9.69.219/home/modules/pdf_modules/m158.pdf)) (PDF file) by Jerzy Borysowicz for Project PHYSNET (<http://www.physnet.org>).

# Zero-point energy

**Zero-point energy** is the lowest possible energy that a quantum mechanical physical system may have; it is the energy of its ground state. All quantum mechanical systems undergo fluctuations even in their ground state and have an associated zero-point energy, a consequence of their wave-like nature. The uncertainty principle requires every physical system to have a zero-point energy greater than the minimum of its classical potential well. This results in motion even at absolute zero. For example, liquid helium does not freeze under atmospheric pressure at any temperature because of its zero-point energy.

The concept of zero-point energy was developed in Germany by Albert Einstein and Otto Stern in 1913, as a corrective term added to a zero-grounded formula developed by Max Planck in 1900.<sup>[1][2]</sup> The term *zero-point energy* originates from the German *Nullpunktsenergie*.<sup>[1][2]</sup> The German name is also spelled *Nullpunktenergie* (without the "s").

Vacuum energy is the zero-point energy of all the fields in space, which in the Standard Model includes the electromagnetic field, other gauge fields, fermionic fields, and the Higgs field. It is the energy of the vacuum, which in quantum field theory is defined not as empty space but as the ground state of the fields. In cosmology, the vacuum energy is one possible explanation for the cosmological constant.<sup>[3]</sup> A related term is *zero-point field*, which is the lowest energy state of a particular field.<sup>[4]</sup>

## History

In 1900, Max Planck derived the formula for the energy of a single *energy radiator*, e.g., a vibrating atomic unit:<sup>[5]</sup>

$$\epsilon = \frac{h\nu}{e^{kT} - 1}$$

where  $h$  is Planck's constant,  $\nu$  is the frequency,  $k$  is Boltzmann's constant, and  $T$  is the absolute temperature.

Then in 1913, using this formula as a basis, Albert Einstein and Otto Stern published a paper of great significance in which they suggested for the first time the existence of a residual energy that all oscillators have at absolute zero. They called this residual energy *Nullpunktsenergie* (German), later translated as *zero-point energy*. They carried out an analysis of the specific heat of hydrogen gas at low temperature, and concluded that the data are best represented if the vibrational energy is<sup>[1][2]</sup>

$$\epsilon = \frac{h\nu}{e^{kT} - 1} + \frac{h\nu}{2}$$

According to this expression, an atomic system at absolute zero retains an energy of  $\frac{1}{2}h\nu$ .

## Relation to the uncertainty principle

Zero-point energy is fundamentally related to the Heisenberg uncertainty principle<sup>[6]</sup>. Roughly speaking, the uncertainty principle states that complementary variables (such as a particle's position and momentum, or a field's value and derivative at a point in space) cannot simultaneously be defined precisely by any given quantum state. In particular, there cannot be a state in which the system sits motionless at the bottom of its potential well, for then its position and momentum would both be completely determined to arbitrarily great precision. Therefore, the lowest-energy state (the ground state) of the system must have a distribution in position and momentum that satisfies the uncertainty principle, which implies its energy must be greater than the minimum of the potential well.

Near the bottom of a potential well, the Hamiltonian of a system (the quantum-mechanical operator giving its energy) can be approximated as

$$\hat{H} = E_0 + \frac{1}{2}k(\hat{x} - x_0)^2 + \frac{1}{2m}\hat{p}^2$$

where  $E_0$  is the minimum of the classical potential well. The uncertainty principle tells us that

$$\sqrt{\langle(\hat{x} - x_0)^2\rangle}\sqrt{\langle\hat{p}^2\rangle} \geq \frac{\hbar}{2},$$

making the expectation values of the kinetic and potential terms above satisfy

$$\left\langle\frac{1}{2}k(\hat{x} - x_0)^2\right\rangle\left\langle\frac{1}{2m}\hat{p}^2\right\rangle \geq \left(\frac{\hbar}{4}\right)^2\frac{k}{m}.$$

The expectation value of the energy must therefore be at least

$$\langle\hat{H}\rangle \geq E_0 + \frac{\hbar}{2}\sqrt{\frac{k}{m}} = E_0 + \frac{\hbar\omega}{2}$$

where  $\omega = \sqrt{k/m}$  is the angular frequency at which the system oscillates.

A more thorough treatment, showing that the energy of the ground state actually is  $E_0 = \hbar\omega/2$ , requires solving for the ground state of the system. See quantum harmonic oscillator for details.

## Varieties

The concept of zero-point energy occurs in a number of situations.

In ordinary quantum mechanics, the zero-point energy is the energy associated with the ground state of the system. The professional physics literature tends to measure frequency, as denoted by  $\nu$  above, using angular frequency, denoted with  $\omega$  and defined by  $\omega = 2\pi\nu$ . This leads to a convention of writing Planck's constant  $\hbar$  with a bar through its top ( $\hbar$ ) to denote the quantity  $h / 2\pi$ . In those terms, the most famous such example of zero-point energy is  $E = \hbar\omega/2$  associated with the ground state of the quantum harmonic oscillator. In quantum mechanical terms, the zero-point energy is the expectation value of the Hamiltonian of the system in the ground state.

In quantum field theory, the fabric of space is visualized as consisting of fields, with the field at every point in space and time being a quantum harmonic oscillator, with neighboring oscillators interacting. In this case, one has a contribution of  $E = \hbar\omega/2$  from every point in space, resulting in a calculation of infinite zero-point energy in any finite volume; this is one reason renormalization is needed to make sense of quantum field theories. The zero-point energy is again the expectation value of the Hamiltonian; here, however, the phrase vacuum expectation value is more commonly used, and the energy is called the vacuum energy.

In quantum perturbation theory, it is sometimes said that the contribution of one-loop and multi-loop Feynman diagrams to elementary particle propagators are the contribution of vacuum fluctuations or the zero-point energy to the particle masses.

## Experimental observations

A phenomenon that is commonly presented as evidence for the existence of zero-point energy in vacuum is the Casimir effect, proposed in 1948 by Dutch physicist Hendrik B. G. Casimir (Philips Research), who considered the quantized electromagnetic field between a pair of grounded, neutral metal plates. The vacuum energy contains contributions from all wavelengths, except those excluded by the spacing between plates. As the plates draw together, more wavelengths are excluded and the vacuum energy decreases. The decrease in energy means there must be a force doing work on the plates as they move. This force has been measured and found to be in good agreement with the theory. However, there is still some debate on whether vacuum energy is necessary to explain the Casimir effect. Robert Jaffe of MIT argues that the Casimir force should not be considered evidence for vacuum

energy, since it can be derived in QED without reference to vacuum energy by considering charge-current interactions (the radiation-reaction picture).<sup>[7]</sup>

The experimentally measured Lamb shift has been argued to be, in part, a zero-point energy effect.<sup>[8]</sup>

## Gravitation and cosmology

In cosmology, the zero-point energy offers an intriguing possibility for explaining the speculative positive values of the proposed cosmological constant. In brief, if the energy is "really there", then it should exert a gravitational force.<sup>[9]</sup> In general relativity, mass and energy are equivalent; both produce a gravitational field. One obvious difficulty with this association is that the zero-point energy of the vacuum is absurdly large. Naively, it is infinite, because it includes the energy of waves with arbitrarily short wavelengths. But since only *differences* in energy are physically measurable, the infinity can be removed by renormalization. In all practical calculations, this is how the infinity is handled. It is also arguable that undiscovered physics relevant at the Planck scale reduces or eliminates the energy of waves shorter than the Planck length, making the total zero-point energy finite.

## Free-energy devices

As a scientific concept, the existence of zero-point energy is not controversial although the ability to harness it is.<sup>[10]</sup> Many people claim to have invented perpetual motion machines and other power-generating devices supposedly based on zero-point energy.

In quantum theory, zero-point energy is a minimum energy below which a thermodynamic system can never go.<sup>[10]</sup> Thus, according to the standard quantum-theoretic viewpoint, none of this energy can be withdrawn without altering the system to a different form in which the system has a lower zero-point energy.<sup>[10]</sup>

Thus, current claims to zero-point-energy-based power generation systems are in contradiction with known physics laws and have the status of pseudoscience.<sup>[10]</sup>

One of the theories that claims that zero-point energy is infinite is stochastic electrodynamics. In it, the zero-point field is viewed as simply a classical background isotropic noise wave field which excites all systems present in the vacuum and thus is *responsible for* their minimum-energy or "ground" states. The requirement of Lorentz invariance at a statistical level then implies that the energy density spectrum must increase with the third power of frequency, implying infinite energy density when integrated over all frequencies.<sup>[11]</sup> However, as pointed out, this energy can't be withdrawn from the system.<sup>[10]</sup>

According to a NASA contractor report, "the concept of accessing a significant amount of useful energy from the ZPE gained much credibility when a major article on this topic was published in *Aviation Week & Space Technology* (March 1st, 2004), a leading aerospace industry magazine".<sup>[12]</sup>

The calculation that underlies the Casimir experiment, a calculation based on the formula predicting infinite vacuum energy, shows the zero-point energy of a system consisting of a vacuum between two plates will decrease at a finite rate as the two plates are drawn together. The vacuum energies are predicted to be infinite, but the changes are predicted to be finite. Casimir combined the projected rate of change in zero-point energy with the principle of conservation of energy to predict a force on the plates. The predicted force, which is very small and was experimentally measured to be within 5% of its predicted value, is finite.<sup>[13]</sup> Even though the zero-point energy is theoretically infinite, there is as yet no practical evidence to suggest that infinite amounts of zero-point energy are available for use, that zero-point energy can be withdrawn for free, or that zero-point energy can be used in violation of conservation of energy.<sup>[14]</sup>

## In popular culture

In Disney/Pixar's animated film "The Incredibles", the main villain Syndrome refers to his weapons as using zero-point energy.<sup>[15]</sup> <sup>[16]</sup> The fan fiction community devoted to the character is named "Zero Point" because of this.<sup>[17]</sup>

In the critically acclaimed game series from Valve Corporation, *Half-Life*, a "zero-point energy field manipulator" (popularly known as 'gravity gun'), meant to handle sensitive, anomalous and hazardous materials, is used as both a weapon to throw objects at enemies in high speeds, as a primary attack, and a tool to solve physics puzzles consisting in moving objects of considerable weight.

In Nintendo's *Star Fox* series, there are mentions of a 'G-Diffusion' system. The so-called 'G-diffusion' is an experimental power system used in the game's starfighters that reduces gravity forces on the pilot and provides a respectable power source for shields and propulsion, using advanced Zero-point energy technology.

Within the *Stargate* franchise the race of beings known as The Ancients who created the *Stargate* devices are also responsible for developing a crystalline energy source that contains a miniature universe from which zero point energy is extracted. Within the canon of both *Stargate SG-1* and *Stargate Atlantis* Zero Point Modules (a.k.a. ZPMs), as they're called, are responsible for producing orders of magnitude greater power than any other known form of energy output in the universe. Three of these were capable of powering the defensive shield on Atlantis, which is roughly the size of Manhattan, while it lay on the ocean floor for over 10,000 years.

## Notes

- [1] Laidler, K. J. (2001). *The World of Physical Chemistry*. Oxford University Press. p. 324. ISBN 0-19-855919-4.
- [2] Einstein, A.; Stern, O. (1913). "Einige Argumente für die Annahme einer molekularen Agitation beim absoluten Nullpunkt". *Annalen der Physik* **40** (3): 551. Bibcode 1913AnP...345..551E. doi:10.1002/andp.19133450309.
- [3] Rugh, S. E.; Zinkernagel, H. (2002). "The Quantum Vacuum and the Cosmological Constant Problem". *Studies in History and Philosophy of Modern Physics*, vol. **33** (4): 663–705. arXiv:hep-th/0012253. doi:10.1016/S1355-2198(02)00033-3.
- [4] Gribbin, J. (1998). *Q is for Quantum: An Encyclopedia of Particle Physics*. Touchstone Books. ISBN 0-684-86315-4.
- [5] Planck, M (1900). "Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum". *Verhandlungen der Deutschen Physikalischen Gesellschaft* **2**: 237–245.
- [6] *Zeitschrift für Physik* **43** (3): 172–198, 1927, doi:10.1007/BF01397280, Original work as HTML (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/bond/papers/corr155.1.html>)
- [7] Jaffe, R. L. (2005). *Physical Review D* **72** (2): 021301.
- [8] Hawton, M. (1993). "Self-consistent frequencies of the electron-photon system". *Physical Review A* **48** (3): 1824. Bibcode 1993PhRvA..48.1824H. doi:10.1103/PhysRevA.48.1824.
- [9] Zee, A. (2008). "Gravity and Its Mysteries: Some Thoughts and Speculations" ([http://www.cospa.ntu.edu.tw/aappsbulletin/data/18-4/06\\_Gravity.pdf](http://www.cospa.ntu.edu.tw/aappsbulletin/data/18-4/06_Gravity.pdf)). *AAPPS Bulletin* **18** (4): 32. .
- [10] "Zero-Point Energy: Can We Get Something From Nothing" (<http://www.scribd.com/doc/297161/Zero-Point-Energy>). U.S. Army National Ground Intelligence Center. . "Forays into "free energy" inventions and perpetual-motion machines using ZPE are considered by the broader scientific community to be pseudoscience."
- [11] de la Peña, L.; Cetto, A. M. (1995). *The Quantum Dice: An Introduction to Stochastic Electrodynamics*. ISBN 978-0-7923-3818-5.
- [12] *Advanced Energetics for Aeronautical Applications: Volume II* ([http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050170447\\_2005172301.pdf](http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050170447_2005172301.pdf)). NASA. .
- [13] "What is the Casimir Effect?" (<http://math.ucr.edu/home/baez/physics/Quantum/casimir.html>). .
- [14] "FOLLOW-UP: What is the 'zero-point energy' (or 'vacuum energy') in quantum physics? Is it really possible that we could harness this energy?" (<http://www.sciam.com/article.cfm?id=follow-up-what-is-the-zer>). *Scientific American*. 18 August 1997. .
- [15] "Buddy Pine character profile" (<http://www.imdb.com/character/ch0003832/>). . Retrieved 5 September 2012.
- [16] "Zero Point Energy" ([http://www.youtube.com/watch?v=\\_wEVBjP3Kk8](http://www.youtube.com/watch?v=_wEVBjP3Kk8)). . Retrieved 5 September 2012.
- [17] "Zero Point at FanFiction.net" ([http://www.fanfiction.net/community/Zero\\_Point/41215/](http://www.fanfiction.net/community/Zero_Point/41215/)). . Retrieved 5 September 2012.

## Bibliography

- Beiser, A. (1967). *Concepts of Modern Physics*. McGraw-Hill. ISBN 0-07-004473-2.
- Einstein, A.; Hopf, L. (1910). "On a theorem of the probability calculus and its application to the theory of radiation". *Annalen der Physik* **33** (16): 1096–1104. Bibcode 1910AnP...338.1096E. doi:10.1002/andp.19103381603.
- Einstein, A.; Hopf, L. (1910). "Statistical investigation of a resonator's motion in a radiation field". *Annalen der Physik* **33** (16): 1105–1115. Bibcode 1910AnP...338.1105E. doi:10.1002/andp.19103381604.
- Einstein, A.; Stern, O. (1913). "Einige Argumente für die Annahme einer molekularen Agitation beim absoluten Nullpunkt". *Annalen der Physik* **40** (3): 551. Bibcode 1913AnP...345..551E. doi:10.1002/andp.19133450309.
- Haisch, B.; Rueda, A.; Dobyns, Y. (2001). "Inertial mass and the quantum vacuum fields" (<http://www.calphysics.org/articles/annalen.pdf>). *Annalen der Physik* **10** (5): 393–414. arXiv:gr-qc/0009036. Bibcode 2001AnP...513..393H. doi:10.1002/1521-3889(200105)10:5<393::AID-ANDP393>3.0.CO;2-Z.
- Loudon, R. (2000). *The Quantum Theory of Light* (3rd ed.). Clarendon Press. ISBN 0-19-850176-5.
- Milonni, P. W. (1994). *The Quantum Vacuum: An Introduction to Quantum Electrodynamics*. Academic Press. ISBN 0-12-498080-5.
- Nernst, W. (1916). "Über einen Versuch, von quantentheoretischen Betrachtungen zur Annahme stetiger Energieänderungen zurückzukehren". *Verhandlungen der Deutschen Physikalischen Gesellschaft* **18**: 83.
- Sciamia, D. W. (1991). Saunders, S.; Brown, H. R.. eds. *The Philosophy of Vacuum*. Clarendon Press. ISBN 0-19-824449-5.
- Rafelski, J.; Muller, B. (1985). *The Structured Vacuum: Thinking about nothing* (<http://www.physics.arizona.edu/~rafelski/Books/StructVacuumE.pdf>). Harri Deutsch. ISBN 3-87144-889-3.
- Boyer, T. H. (1968). "Quantum Electromagnetic Zero-Point Energy of a Conducting Spherical Shell and the Casimir Model for a Charged Particle" (<http://www.mit.edu/~kardar/research/seminars/Casimir/PR-Boyer68.pdf>). *Physical Review* **174** (5): 13. Bibcode 1968PhRv..174.1764B. doi:10.1103/PhysRev.174.1764.
- Yablonovitch, E. (1989). "Accelerating Reference Frame for Electromagnetic Waves in a Rapidly Growing Plasma: Unruh-Davies-Fulling-DeWitt Radiation and the Nonadiabatic Casimir Effect" (<http://www.ee.ucla.edu/~photon/pubs/ey1989prl6215.pdf>). *Physical Review Letters* **62** (15): 1742–1745. Bibcode 1989PhRvL..62.1742Y. doi:10.1103/PhysRevLett.62.1742. PMID 10039756.

## External links

- Zero-point energy? (<http://van.physics.uiuc.edu/qa/listing.php?id=1256>) "Ask the Van" popular science FAQ at University of Illinois.
- Philip Yam, "Exploiting Zero-point Energy" (<http://www.padrak.com/ine/ZPESCIAM.html>), *Scientific American Magazine*, December 1997, pp. 82–85.
- Matt Visser, What is the 'zero-point energy' (or 'vacuum energy') in quantum physics? Is it really possible that we could harness this energy? (<http://homepages.mcs.vuw.ac.nz/~visser/general.shtml#what-zpe>) from *Scientific American Magazine*, August 18, 1997 (<http://www.sciam.com/article.cfm?id=follow-up-what-is-the-zer>)
- Calphysics Institute "Zero Point Energy and Zero Point Field" (<http://www.calphysics.org/zpe.html>)

# Vacuum energy

---

**Vacuum energy** is an underlying background energy that exists in space throughout the entire Universe. One contribution to the vacuum energy may be from virtual particles, which are thought to be particle pairs that blink into existence and then annihilate in a timespan too short to observe. They are expected to do this everywhere, throughout the Universe. Their behavior is codified in Heisenberg's energy-time uncertainty principle. Still, the exact effect of such fleeting bits of energy is difficult to quantify.

The effects of vacuum energy can be experimentally observed in various phenomena such as spontaneous emission, the Casimir effect and the Lamb shift, and are thought to influence the behavior of the Universe on cosmological scales. Using the upper limit of the cosmological constant, the vacuum energy in a cubic meter of free space has been estimated to be  $10^{-9}$  Joules.<sup>[1]</sup> However, in both Quantum Electrodynamics (QED) and Stochastic Electrodynamics (SED), consistency with the principle of Lorentz covariance and with the magnitude of the Planck Constant requires it to have a much larger value of  $10^{113}$  Joules per cubic meter.<sup>[2][3]</sup> See vacuum catastrophe.

## Origin

Quantum field theory states that all fundamental fields, such as the electromagnetic field, must be quantized at each and every point in space. A field in physics may be envisioned as if space were filled with interconnected vibrating balls and springs, and the strength of the field were like the displacement of a ball from its rest position. The theory requires "vibrations" in, or more accurately changes in the strength of, such a field to propagate as per the appropriate wave equation for the particular field in question. The second quantization of quantum field theory requires that each such ball-spring combination be quantized, that is, that the strength of the field be quantized at each point in space. Canonically, if the field at each point in space is a simple harmonic oscillator, its quantization places a quantum harmonic oscillator at each point. Excitations of the field correspond to the elementary particles of particle physics. Thus, according to the theory, even the vacuum has a vastly complex structure and all calculations of quantum field theory must be made in relation to this model of the vacuum.

The theory considers vacuum to implicitly have the same properties as a particle, such as spin or polarization in the case of light, energy, and so on. According to the theory, most of these properties cancel out on average leaving the vacuum empty in the literal sense of the word. One important exception, however, is the vacuum energy or the vacuum expectation value of the energy. The quantization of a simple harmonic oscillator requires the lowest possible energy, or zero-point energy of such an oscillator to be:

$$E = \frac{1}{2}h\nu.$$

Summing over all possible oscillators at all points in space gives an infinite quantity. To remove this infinity, one may argue that only differences in energy are physically measurable, much as the concept of potential energy has been treated in classical mechanics for centuries. This argument is the underpinning of the theory of renormalization. In all practical calculations, this is how the infinity is handled.

Vacuum energy can also be thought of in terms of virtual particles (also known as vacuum fluctuations) which are created and destroyed out of the vacuum. These particles are always created out of the vacuum in particle-antiparticle pairs, which in most cases shortly annihilate each other and disappear. However, these particles and antiparticles may interact with others before disappearing, a process which can be mapped using Feynman diagrams. Note that this method of computing vacuum energy is mathematically equivalent to having a quantum harmonic oscillator at each point and, therefore, suffers the same renormalization problems.

Additional contributions to the vacuum energy come from spontaneous symmetry breaking in quantum field theory.

---

## Implications

Vacuum energy has a number of consequences. In 1948, Dutch physicists Hendrik B. G. Casimir and Dirk Polder predicted the existence of a tiny attractive force between closely placed metal plates due to resonances in the vacuum energy in the space between them. This is now known as the Casimir effect and has since been extensively experimentally verified. It is therefore believed that the vacuum energy is "real" in the same sense that more familiar conceptual objects such as electrons, magnetic fields, etc., are real. However, alternate explanations for the Casimir have since been proposed.<sup>[4]</sup>

Other predictions are harder to verify. Vacuum fluctuations are always created as particle/antiparticle pairs. The creation of these virtual particles near the event horizon of a black hole has been hypothesized by physicist Stephen Hawking to be a mechanism for the eventual "evaporation" of black holes. The net energy of the Universe remains zero so long as the particle pairs annihilate each other within Planck time. If one of the pair is pulled into the black hole before this, then the other particle becomes "real" and energy/mass is essentially radiated into space from the black hole. This loss is cumulative and could result in the black hole's disappearance over time. The time required is dependent on the mass of the black hole but could be on the order of  $10^{100}$  years for large solar-mass black holes.

The vacuum energy also has important consequences for physical cosmology. Special relativity predicts that energy is equivalent to mass, and therefore, if the vacuum energy is "really there", it should exert a gravitational force. Essentially, a non-zero vacuum energy is expected to contribute to the cosmological constant, which affects the expansion of the universe. In the special case of vacuum energy, general relativity stipulates that the gravitational field is proportional to  $\rho - 3p$  (where  $\rho$  is the mass-energy density, and  $p$  is the pressure). Quantum theory of the vacuum further stipulates that the pressure of the zero-state vacuum energy is always negative and equal to  $\rho$ . Thus, the total of  $\rho - 3p$  becomes  $-2\rho$ : A negative value. This calculation implies a repulsive gravitational field, giving rise to expansion, if indeed the vacuum ground state has non-zero energy. However, the vacuum energy is mathematically infinite without renormalization, which is based on the assumption that we can only measure energy in a relative sense, which is not true if we can observe it indirectly via the cosmological constant.

The existence of vacuum energy is also sometimes used as theoretical justification for the possibility of free energy machines. It has been argued that due to the broken symmetry (in QED), free energy does not violate conservation of energy, since the laws of thermodynamics only apply to equilibrium systems. However, consensus amongst physicists is that this is incorrect and that vacuum energy cannot be harnessed to generate free energy.<sup>[5]</sup> In particular, the second law of thermodynamics is unaffected by the existence of vacuum energy. However, in Stochastic Electrodynamics, the energy density is taken to be a classical random noise wave field which consists of real electromagnetic noise waves propagating isotropically in all directions. The energy in such a wave field would seem to be accessible, e.g., with nothing more complicated than a directional coupler. The most obvious difficulty appears to be the spectral distribution of the energy, which compatibility with Lorentz invariance requires to take the form  $Kf^3$ , where  $K$  is a constant and  $f$  denotes frequency.<sup>[6][7]</sup> It follows that the energy and momentum flux in this wave field only becomes significant at extremely short wavelengths where directional coupler technology is currently lacking.

## History

In 1934, Georges Lemaître used an unusual perfect-fluid equation of state to interpret the cosmological constant as due to vacuum energy. In 1948, the Casimir effect provided the first experimental verification of the existence of vacuum energy. In 1957, Lee and Yang proved the concepts of broken symmetry and parity violation, for which they won the Nobel prize. In 1973, Edward Tryon proposed that the Universe may be a large-scale quantum-mechanical vacuum fluctuation where positive mass-energy is balanced by negative gravitational potential energy. During the 1980s, there were many attempts to relate the fields that generate the vacuum energy to specific fields that were predicted by attempts at a Grand unification theory and to use observations of the Universe to confirm one or another version. However, the exact nature of the particles (or fields) that generate vacuum energy, with a density such as

that required by inflation theory, remains a mystery.

## External articles and references

- Free pdf copy of *The Structured Vacuum - thinking about nothing* <sup>[8]</sup> by Johann Rafelski and Berndt Muller (1985) ISBN 3-87144-889-3.
- Saunders, S., & Brown, H. R. (1991). *The Philosophy of Vacuum*. Oxford [England]: Clarendon Press.
- Poincaré Seminar, Duplantier, B., & Rivasseau, V. (2003). "Poincaré Seminar 2002: vacuum energy-renormalization". *Progress in mathematical physics*, v. 30. Basel: Birkhäuser Verlag.
- Futamase & Yoshida *Possible measurement of vacuum energy* <sup>[9]</sup>
- YAN Kun (2006), Vacuum energy and superluminal velocity <sup>[10]</sup>
- G. Jordan Maclay, (et al.): *Study of Vacuum Energy Physics for Breakthrough Propulsion*. <sup>[11]</sup> 2004, NASA Glenn Technical Reports Server, (pdf, 57 pages, Retrieved 2009-10-22)

## Notes

- [1] Sean Carroll, Sr Research Associate - Physics, California Institute of Technology, June 22, 2006C-SPAN broadcast of Cosmology at Yearly Kos Science Panel, Part 1
- [2] Peter W. Milonni - "The Quantum Vacuum"
- [3] de la Pena and Cetto "The Quantum Dice: An Introduction to Stochastic Electrodynamics"
- [4] R. L. Jaffe: *The Casimir Effect and the Quantum Vacuum*. In: *Physical Review D*. Band 72, 2005 (<http://arxiv.org/abs/hep-th/0503158>)
- [5] IEEE Trans. Ed., 1996, p.7 ([http://www.eleceng.adelaide.edu.au/Personal/dabbott/publications/EDU\\_abbott1996.pdf](http://www.eleceng.adelaide.edu.au/Personal/dabbott/publications/EDU_abbott1996.pdf))
- [6] Peter W. Milonni - "The Quantum Vacuum"
- [7] de la Pena and Cetto "The Quantum Dice: An Introduction to Stochastic Electrodynamics"
- [8] <http://www.physics.arizona.edu/~rafelski/Books/StructVacuumE.pdf>
- [9] <http://arxiv.org/abs/gr-qc/0011083v1>
- [10] <http://www.slac.stanford.edu/spires/find/hep/www?irn=5820570>
- [11] <http://gltrs.grc.nasa.gov/reports/2004/CR-2004-213311.pdf>

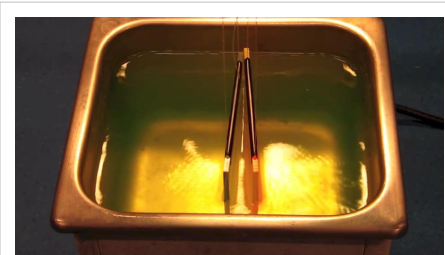
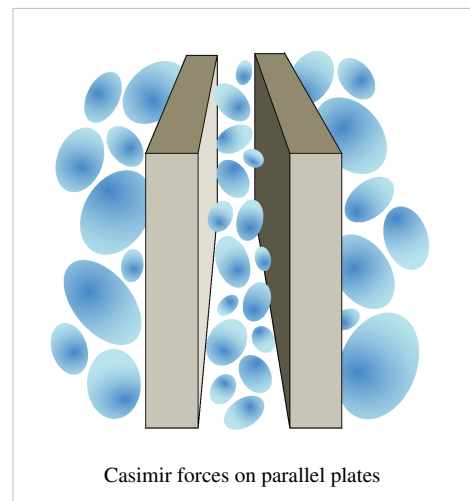
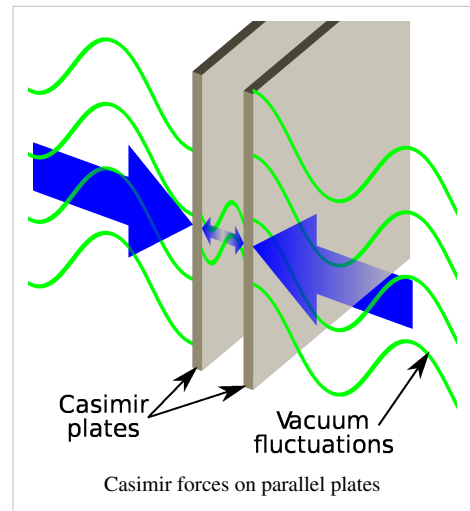


# Casimir effect

In quantum field theory, the **Casimir effect** and the **Casimir–Polder force** are physical forces arising from a quantized field. The typical example is of two uncharged metallic plates in a vacuum, placed a few micrometers apart as in a capacitor but without any external electromagnetic field. In a classical description, the lack of an external field also means that there is no field between the plates, and no force would be measured between them.<sup>[1]</sup> When this field is instead studied using the QED vacuum of quantum electrodynamics, it is seen that the plates do affect the virtual photons which constitute the field, and generate a net force<sup>[2]</sup>—either an attraction or a repulsion depending on the specific arrangement of the two plates. Although the Casimir effect can be expressed in terms of virtual particles interacting with the objects, it is best described and more easily calculated in terms of the zero-point energy of a quantized field in the intervening space between the objects. This force has been measured, and is a striking example of an effect purely due to second quantization.<sup>[3][4]</sup> However, the treatment of boundary conditions in these calculations has led to some controversy. In fact "Casimir's original goal was to compute the van der Waals force between polarizable molecules" of the metallic plates. Thus it can be interpreted without any reference to the zero-point energy (vacuum energy) or virtual particles of quantum fields.<sup>[5]</sup>

Dutch physicists Hendrik B. G. Casimir and Dirk Polder at Philips Research Labs proposed the existence of a force between two polarizable atoms and between such an atom and a conducting plate in 1947, and, after a conversation with Niels Bohr who suggested it had something to do with zero-point energy, Casimir alone formulated the theory predicting a force between neutral conducting plates in 1948; the former is called the Casimir-Polder force while the latter is the Casimir effect in the narrow sense. Predictions of the force were later extended to finite-conductivity metals and dielectrics by Lifshitz and his students, and recent calculations have considered more general geometries. It was not until 1997, however, that a direct experiment, by S. Lamoreaux, described above, quantitatively measured the force (to within 15% of the value predicted by the theory),<sup>[6]</sup> although previous work [e.g. van Blokland and Overbeek (1978)] had observed the force qualitatively, and indirect validation of the predicted Casimir energy had been made by measuring the thickness of liquid Helium films by Sabisky and Anderson in 1972. Subsequent experiments approach an accuracy of a few percent.

Because the strength of the force falls off rapidly with distance, it is only measurable when the distance between the objects is extremely small. On a submicron scale, this force becomes so strong that it becomes the dominant force between uncharged conductors. In fact, at separations of 10 nm—about 100 times the typical size of an atom—the Casimir effect produces the equivalent of 1 atmosphere of pressure (101.325 kPa), the precise value depending on



A water wave analogue of the Casimir effect. Two parallel plates are submerged into colored water contained in a sonicator. When the sonicator is turned on, waves are excited imitating vacuum fluctuations; as a result, the plates attract to each other.

surface geometry and other factors.<sup>[7]</sup>

In modern theoretical physics, the Casimir effect plays an important role in the chiral bag model of the nucleon; and in applied physics, it is significant in some aspects of emerging microtechnologies and nanotechnologies.<sup>[8]</sup>

Any medium supporting oscillations has an analogue of the Casimir effect. For example, beads on a string<sup>[9][10]</sup> as well as plates submerged in noisy water<sup>[11]</sup> or gas<sup>[12]</sup> exhibit the Casimir force.

## Overview

The Casimir effect can be understood by the idea that the presence of conducting metals and dielectrics alters the vacuum expectation value of the energy of the second quantized electromagnetic field.<sup>[13][14]</sup> Since the value of this energy depends on the shapes and positions of the conductors and dielectrics, the Casimir effect manifests itself as a force between such objects.

## Possible causes

### Vacuum energy

The causes of the Casimir effect are described by quantum field theory, which states that all of the various fundamental fields, such as the electromagnetic field, must be quantized at each and every point in space. In a simplified view, a "field" in physics may be envisioned as if space were filled with interconnected vibrating balls and springs, and the strength of the field can be visualized as the displacement of a ball from its rest position. Vibrations in this field propagate and are governed by the appropriate wave equation for the particular field in question. The second quantization of quantum field theory requires that each such ball-spring combination be quantized, that is, that the strength of the field be quantized at each point in space. At the most basic level, the field at each point in space is a simple harmonic oscillator, and its quantization places a quantum harmonic oscillator at each point. Excitations of the field correspond to the elementary particles of particle physics. However, even the vacuum has a vastly complex structure, so all calculations of quantum field theory must be made in relation to this model of the vacuum.

The vacuum has, implicitly, all of the properties that a particle may have: spin, or polarization in the case of light, energy, and so on. On average, most of these properties cancel out: the vacuum is, after all, "empty" in this sense. One important exception is the vacuum energy or the vacuum expectation value of the energy. The quantization of a simple harmonic oscillator states that the lowest possible energy or zero-point energy that such an oscillator may have is

$$E = \frac{1}{2} \hbar \omega .$$

Summing over all possible oscillators at all points in space gives an infinite quantity. To remove this infinity, one may argue that only differences in energy are physically measurable; this argument is the underpinning of the theory of renormalization. In all practical calculations, this is how the infinity is always handled. In a deeper sense, however, renormalization is unsatisfying, and the removal of this infinity presents a challenge in the search for a Theory of Everything. Currently there is no compelling explanation for how this infinity should be treated as essentially zero; a non-zero value is essentially the cosmological constant and any large value causes trouble in cosmology.

## Relativistic van der Waals force

Alternatively, a 2005 paper by Robert Jaffe of MIT states that "Casimir effects can be formulated and Casimir forces can be computed without reference to zero point energies. They are relativistic, quantum forces between charges and currents. The Casimir force (per unit area) between parallel plates vanishes as  $\alpha$ , the fine structure constant, goes to zero, and the standard result, which appears to be independent of  $\alpha$ , corresponds to the  $\alpha \rightarrow \infty$  limit," and that "The Casimir force is simply the (relativistic, retarded) van der Waals force between the metal plates."<sup>[15]</sup>

## Effects

Casimir's observation was that the second-quantized quantum electromagnetic field, in the presence of bulk bodies such as metals or dielectrics, must obey the same boundary conditions that the classical electromagnetic field must obey. In particular, this affects the calculation of the vacuum energy in the presence of a conductor or dielectric.

Consider, for example, the calculation of the vacuum expectation value of the electromagnetic field inside a metal cavity, such as, for example, a radar cavity or a microwave waveguide. In this case, the correct way to find the zero point energy of the field is to sum the energies of the standing waves of the cavity. To each and every possible standing wave corresponds an energy; say the energy of the  $n$ th standing wave is  $E_n$ . The vacuum expectation value of the energy of the electromagnetic field in the cavity is then

$$\langle E \rangle = \frac{1}{2} \sum_n E_n$$

with the sum running over all possible values of  $n$  enumerating the standing waves. The factor of 1/2 corresponds to the fact that the zero-point energies are being summed (it is the same 1/2 as appears in the equation  $E = \hbar\omega/2$ ).

Written in this way, this sum is clearly divergent; however, it can be used to create finite expressions.

In particular, one may ask how the zero point energy depends on the shape  $s$  of the cavity. Each energy level  $E_n$  depends on the shape, and so one should write  $E_n(s)$  for the energy level, and  $\langle E(s) \rangle$  for the vacuum expectation value. At this point comes an important observation: the force at point  $p$  on the wall of the cavity is equal to the change in the vacuum energy if the shape  $s$  of the wall is perturbed a little bit, say by  $\delta s$ , at point  $p$ . That is, one has

$$F(p) = - \left. \frac{\delta \langle E(s) \rangle}{\delta s} \right|_p.$$

This value is finite in many practical calculations.<sup>[16]</sup>

Attraction between the plates can be easily understood by focusing on the 1-dimensional situation. Suppose that a moveable conductive plate is positioned at a short distance  $a$  from one of two widely separated plates (distance  $L$  apart). With  $a \ll L$ , the states within the slot of width  $a$  are highly constrained so that the energy  $E$  of any one mode is widely separated from that of the next. This is not the case in open region  $L$ , where there is a large number (about  $L/a$ ) of states with energy evenly spaced between  $E$  and the next mode in the narrow slot--in other words, all slightly larger than  $E$ . Now on shortening  $a$  by  $da$  ( $< 0$ ), the mode in the slot shrinks in wavelength and therefore increases in energy proportional to  $-da/a$ , whereas all the outside  $L/a$  states lengthen and correspondingly lower energy proportional to  $da/L$  (note the denominator). The net change is slightly negative, because all the  $L/a$  modes' energies are slightly larger than the single mode in the slot.

## Derivation of Casimir effect assuming zeta-regularization

In the original calculation done by Casimir, he considered the space between a pair of conducting metal plates at distance  $a$  apart. In this case, the standing waves are particularly easy to calculate, since the transverse component of the electric field and the normal component of the magnetic field must vanish on the surface of a conductor. Assuming the parallel plates lie in the  $xy$ -plane, the standing waves are

$$\psi_n(x, y, z; t) = e^{-i\omega_n t} e^{ik_x x + ik_y y} \sin(k_n z)$$

where  $\psi$  stands for the electric component of the electromagnetic field, and, for brevity, the polarization and the magnetic components are ignored here. Here,  $k_x$  and  $k_y$  are the wave vectors in directions parallel to the plates, and

$$k_n = \frac{n\pi}{a}$$

is the wave-vector perpendicular to the plates. Here,  $n$  is an integer, resulting from the requirement that  $\psi$  vanish on the metal plates. The energy of this wave is

$$\omega_n = c \sqrt{k_x^2 + k_y^2 + \frac{n^2 \pi^2}{a^2}}$$

where  $c$  is the speed of light. The vacuum energy is then the sum over all possible excitation modes

$$\langle E \rangle = \frac{\hbar}{2} \cdot 2 \int \frac{A dk_x dk_y}{(2\pi)^2} \sum_{n=1}^{\infty} \omega_n$$

where  $A$  is the area of the metal plates, and a factor of 2 is introduced for the two possible polarizations of the wave. This expression is clearly infinite, and to proceed with the calculation, it is convenient to introduce a regulator (discussed in greater detail below). The regulator will serve to make the expression finite, and in the end will be removed. The zeta-regularized version of the energy per unit-area of the plate is

$$\frac{\langle E(s) \rangle}{A} = \hbar \int \frac{dk_x dk_y}{(2\pi)^2} \sum_{n=1}^{\infty} \omega_n |\omega_n|^{-s}.$$

In the end, the limit  $s \rightarrow 0$  is to be taken. Here  $s$  is just a complex number, not to be confused with the shape discussed previously. This integral/sum is finite for  $s$  real and larger than 3. The sum has a pole at  $s = 3$ , but may be analytically continued to  $s = 0$ , where the expression is finite. The above expression simplifies to:

$$\frac{\langle E(s) \rangle}{A} = \frac{\hbar c^{1-s}}{4\pi^2} \sum_n \int_0^\infty 2\pi q dq \left| q^2 + \frac{\pi^2 n^2}{a^2} \right|^{(1-s)/2}$$

where polar coordinates  $q^2 = k_x^2 + k_y^2$  were introduced to turn the double integral into a single integral. The  $q$  in front is the Jacobian, and the  $2\pi$  comes from the angular integration. The integral converges if  $\text{Re}[s] > 3$ , resulting in

$$\frac{\langle E(s) \rangle}{A} = -\frac{\hbar c^{1-s} \pi^{2-s}}{2a^{3-s}} \frac{1}{3-s} \sum_n |n|^{3-s}.$$

The sum diverges at  $s$  in the neighborhood of zero, but if the damping of large-frequency excitations corresponding to analytic continuation of the Riemann zeta function to  $s = 0$  is assumed to make sense physically in some way, then one has

$$\frac{\langle E \rangle}{A} = \lim_{s \rightarrow 0} \frac{\langle E(s) \rangle}{A} = -\frac{\hbar c \pi^2}{6a^3} \zeta(-3).$$

But  $\zeta(-3) = 1/120$  and so one obtains

$$\frac{\langle E \rangle}{A} = \frac{-\hbar c \pi^2}{3 \cdot 240 a^3}.$$

The analytic continuation has evidently lost an additive positive infinity, somehow exactly accounting for the zero-point energy (not included above) outside the slot between the plates, but which changes upon plate movement within a closed system. The Casimir force per unit area  $F_c/A$  for idealized, perfectly conducting plates with vacuum between them is

$$\frac{F_c}{A} = -\frac{d}{da} \frac{\langle E \rangle}{A} = -\frac{\hbar c \pi^2}{240 a^4}$$

where

$\hbar$  (hbar,  $\hbar$ ) is the reduced Planck constant,

$c$  is the speed of light,

$a$  is the distance between the two plates.

The force is negative, indicating that the force is attractive: by moving the two plates closer together, the energy is lowered. The presence of  $\hbar$  shows that the Casimir force per unit area  $F_c/A$  is very small, and that furthermore, the force is inherently of quantum-mechanical origin.

NOTE: In Casimir's original derivation [17], a moveable conductive plate is positioned at a short distance  $a$  from one of two widely separated plates (distance  $L$  apart). The 0-point energy on *both* sides of the plate is considered. Instead of the above *ad hoc* analytic continuation assumption, non-convergent sums and integrals are computed using Euler-Maclaurin summation with a regularizing function (e.g., exponential regularization) not so anomalous as  $|\omega_n|^{-s}$  in the above.

### More recent theory

Casimir's analysis of idealized metal plates was generalized to arbitrary dielectric and realistic metal plates by Lifshitz and his students.<sup>[18][19]</sup> Using this approach, complications of the bounding surfaces, such as the modifications to the Casimir force due to finite conductivity, can be calculated numerically using the tabulated complex dielectric functions of the bounding materials. Lifshitz' theory for two metal plates reduces to Casimir's idealized  $1/a^4$  force law for large separations  $a$  much greater than the skin depth of the metal, and conversely reduces to the  $1/a^3$  force law of the London dispersion force (with a coefficient called a Hamaker constant) for small  $a$ , with a more complicated dependence on  $a$  for intermediate separations determined by the dispersion of the materials.<sup>[20]</sup>

Lifshitz' result was subsequently generalized to arbitrary multilayer planar geometries as well as to anisotropic and magnetic materials, but for several decades the calculation of Casimir forces for non-planar geometries remained limited to a few idealized cases admitting analytical solutions.<sup>[21]</sup> For example, the force in the experimental sphere–plate geometry was computed with an approximation (due to Derjaguin) that the sphere radius  $R$  is much larger than the separation  $a$ , in which case the nearby surfaces are nearly parallel and the parallel-plate result can be adapted to obtain an approximate  $R/a^3$  force (neglecting both skin-depth and higher-order curvature effects).<sup>[21][22]</sup> However, in the 2000s a number of authors developed and demonstrated a variety of numerical techniques, in many cases adapted from classical computational electromagnetics, that are capable of accurately calculating Casimir forces for arbitrary geometries and materials, from simple finite-size effects of finite plates to more complicated phenomena arising for patterned surfaces or objects of various shapes.<sup>[21]</sup>

## Measurement

One of the first experimental tests was conducted by Marcus Sparnaay at Philips in Eindhoven, in 1958, in a delicate and difficult experiment with parallel plates, obtaining results not in contradiction with the Casimir theory,<sup>[23][24]</sup> but with large experimental errors. Some of the experimental details as well as some background information on how Casimir, Polder and Sparnaay arrived at this point<sup>[25]</sup> are highlighted in a 2007 interview with Marcus Sparnaay.

The Casimir effect was measured more accurately in 1997 by Steve K. Lamoreaux of Los Alamos National Laboratory,<sup>[26]</sup> and by Umar Mohideen and Anushree Roy of the University of California at Riverside.<sup>[27]</sup> In practice, rather than using two parallel plates, which would require phenomenally accurate alignment to ensure they were parallel, the experiments use one plate that is flat and another plate that is a part of a sphere with a large radius.

In 2001, a group (Giacomo Bressi, Gianni Carugno, Roberto Onofrio and Giuseppe Ruoso) at the University of Padua (Italy) finally succeeded in measuring the Casimir force between parallel plates using microresonators.<sup>[28]</sup>

## Regularisation

In order to be able to perform calculations in the general case, it is convenient to introduce a regulator in the summations. This is an artificial device, used to make the sums finite so that they can be more easily manipulated, followed by the taking of a limit so as to remove the regulator.

The heat kernel or exponentially regulated sum is

$$\langle E(t) \rangle = \frac{1}{2} \sum_n \hbar |\omega_n| \exp(-t|\omega_n|)$$

where the limit  $t \rightarrow 0^+$  is taken in the end. The divergence of the sum is typically manifested as

$$\langle E(t) \rangle = \frac{C}{t^3} + \text{finite}$$

for three-dimensional cavities. The infinite part of the sum is associated with the bulk constant  $C$  which *does not* depend on the shape of the cavity. The interesting part of the sum is the finite part, which is shape-dependent. The Gaussian regulator

$$\langle E(t) \rangle = \frac{1}{2} \sum_n \hbar |\omega_n| \exp(-t^2|\omega_n|^2)$$

is better suited to numerical calculations because of its superior convergence properties, but is more difficult to use in theoretical calculations. Other, suitably smooth, regulators may be used as well. The zeta function regulator

$$\langle E(s) \rangle = \frac{1}{2} \sum_n \hbar |\omega_n| |\omega_n|^{-s}$$

is completely unsuited for numerical calculations, but is quite useful in theoretical calculations. In particular, divergences show up as poles in the complex  $s$  plane, with the bulk divergence at  $s = 4$ . This sum may be analytically continued past this pole, to obtain a finite part at  $s = 0$ .

Not every cavity configuration necessarily leads to a finite part (the lack of a pole at  $s = 0$ ) or shape-independent infinite parts. In this case, it should be understood that additional physics has to be taken into account. In particular, at extremely large frequencies (above the plasma frequency), metals become transparent to photons (such as X-rays), and dielectrics show a frequency-dependent cutoff as well. This frequency dependence acts as a natural regulator. There are a variety of bulk effects in solid state physics, mathematically very similar to the Casimir effect, where the cutoff frequency comes into explicit play to keep expressions finite. (These are discussed in greater detail in *Landau and Lifshitz*, "Theory of Continuous Media".)

## Generalities

The Casimir effect can also be computed using the mathematical mechanisms of functional integrals of quantum field theory, although such calculations are considerably more abstract, and thus difficult to comprehend. In addition, they can be carried out only for the simplest of geometries. However, the formalism of quantum field theory makes it clear that the vacuum expectation value summations are in a certain sense summations over so-called "virtual particles".

More interesting is the understanding that the sums over the energies of standing waves should be formally understood as sums over the eigenvalues of a Hamiltonian. This allows atomic and molecular effects, such as the van der Waals force, to be understood as a variation on the theme of the Casimir effect. Thus one considers the Hamiltonian of a system as a function of the arrangement of objects, such as atoms, in configuration space. The change in the zero-point energy as a function of changes of the configuration can be understood to result in forces acting between the objects.

In the chiral bag model of the nucleon, the Casimir energy plays an important role in showing the mass of the nucleon is independent of the bag radius. In addition, the spectral asymmetry is interpreted as a non-zero vacuum expectation value of the baryon number, cancelling the topological winding number of the pion field surrounding the nucleon.

## Dynamical Casimir effect

The dynamical Casimir effect is the production of particles and energy from a very fast *moving mirror*. This reaction was predicted by certain numerical solutions to quantum mechanics equations made in the 1970's.<sup>[29]</sup> In May 2011 an announcement was made by researchers at the Chalmers University of Technology, in Gothenburg, Sweden, of the detection of the dynamical Casimir effect. These researchers used a modified SQUID to provide the mirror moving at the required relativistic velocity. If confirmed this would be the first experimental verification of the dynamical Casimir effect.<sup>[30] [31]</sup>

## Analogies

A similar analysis can be used to explain Hawking radiation that causes the slow "evaporation" of black holes (although this is generally visualized as the escape of one particle from a virtual particle-antiparticle pair, the other particle having been captured by the black hole).

Constructed within the framework of quantum field theory in curved spacetime, the dynamical Casimir effect has been used to better understand acceleration radiation such as the Unruh effect.

## Repulsive forces

There are few instances wherein the Casimir effect can give rise to repulsive forces between uncharged objects. In a seminal paper, Evgeny Lifshitz showed (theoretically) that in certain circumstances (most commonly involving liquids), repulsive forces can arise.<sup>[32]</sup> This has sparked interest in applications of the Casimir effect toward the development of levitating devices. An experimental demonstration of the Casimir-based repulsion predicted by Lifshitz was recently carried out by Munday et al.<sup>[33]</sup> Other scientists have also suggested the use of gain media to achieve a similar levitation effect,<sup>[34]</sup> though this is controversial because these materials seem to violate fundamental causality constraints and the requirement of thermodynamic equilibrium (Kramers-Kronig relations). Casimir and Casimir-Polder repulsion can in fact occur for sufficiently anisotropic electrical bodies; for a review of the issues involved with repulsion see Milton et al.<sup>[35]</sup>

## Applications

It has been suggested that the Casimir forces have application in nanotechnology,<sup>[36]</sup> in particular silicon integrated circuit technology based micro- and nanoelectromechanical systems, silicon array propulsion for space drives, and so-called Casimir oscillators.<sup>[37]</sup>

## References

- [1] Cyriaque Genet, Francesco Intravaia, Astrid Lambrecht and Serge Reynaud (2004) "Electromagnetic vacuum fluctuations, Casimir and Van der Waals forces ([http://arxiv.org/PS\\_cache/quant-ph/pdf/0302/0302072v2.pdf](http://arxiv.org/PS_cache/quant-ph/pdf/0302/0302072v2.pdf))"
- [2] The Force of Empty Space (<http://focus.aps.org/story/v2/st28>), Physical Review Focus, 3 December 1998
- [3] A. Lambrecht, The Casimir effect: a force from nothing (<http://physicsworld.com/cws/article/print/9747>), *Physics World*, September 2002.
- [4] American Institute of Physics News Note 1996 (<http://www.aip.org/pnu/1996/split/pnu300-3.htm>)
- [5] Jaffe, R. (2005). "Casimir effect and the quantum vacuum". *Physical Review D* **72** (2): 021301. arXiv:hep-th/0503158. Bibcode 2005PhRvD..72b1301J. doi:10.1103/PhysRevD.72.021301.
- [6] Photo of ball attracted to a plate by Casimir effect (<http://apod.nasa.gov/apod/ap061217.html>)
- [7] "The Casimir effect: a force from nothing" (<http://physicsworld.com/cws/article/print/9747>). physicsworld.com. 1 September 2002. . Retrieved 17 July 2009.
- [8] Astrid Lambrecht, Serge Reynaud and Cyriaque Genet" Casimir In The Nanoworld (<http://www.fisica.unipa.it/~cewqo2007/Archive/presentations/Genet.pdf>)"
- [9] Griffiths, D. J.; Ho, E. (2001). "Classical Casimir effect for beads on a string". *American Journal of Physics* **69** (11): 1173. doi:10.1119/1.1396620.
- [10] Cooke, J. H. (1998). "Casimir force on a loaded string". *American Journal of Physics* **66** (7): 569–561. doi:10.1119/1.18907.
- [11] Denardo, B. C.; Puda, J. J.; Larraza, A. S. (2009). "A water wave analog of the Casimir effect". *American Journal of Physics* **77** (12): 1095. doi:10.1119/1.3211416.
- [12] Larraza, A. S.; Denardo, B. (1998). "An acoustic Casimir effect". *Physics Letters A* **248** (2–4): 151. doi:10.1016/S0375-9601(98)00652-5.
- [13] E. L. Losada" Functional Approach to the Fermionic Casimir Effect (<http://particulas.cnea.gov.ar/workshops/silafae/data/226.pdf>)"
- [14] Michael Bordag, Galina Leonidovna Klimchitskaya, Umar Mohideen (2009). "Chapter I; §3: Field quantization and vacuum energy in the presence of boundaries" ([http://books.google.com/books?id=CqE1f\\_s5PgYC&pg=PA33](http://books.google.com/books?id=CqE1f_s5PgYC&pg=PA33)). *Advances in the Casimir effect*. Oxford University Press. pp. 33 ff. ISBN 0-19-923874-X. .
- [15] R.L.Jaffe (2005). "The Casimir Effect and the Quantum Vacuum" (<http://arxiv.org/pdf/hep-th/0503158v1.pdf>). ArXiv preprint. .
- [16] For a brief summary, see the introduction in Passante, R.; Spagnolo, S. (2007). "Casimir-Polder interatomic potential between two atoms at finite temperature and in the presence of boundary conditions". *Physical Review A* **76** (4): 042112. arXiv:0708.2240. Bibcode 2007PhRvA..76d2112P. doi:10.1103/PhysRevA.76.042112.
- [17] <http://www.dwc.knaw.nl/DL/publications/PU00018547.pdf>
- [18] Dzyaloshinskii, I E; Lifshitz, E M; Pitaevskii, Lev P (1961). "GENERAL THEORY OF VAN DER WAALS' FORCES". *Soviet Physics Uspekhi* **4** (2): 153. Bibcode 1961SvPhU...4..153D. doi:10.1070/PU1961v004n02ABEH003330.
- [19] Dzyaloshinskii, I E; Kats, E I (2004). "Casimir forces in modulated systems". *Journal of Physics: Condensed Matter* **16** (32): 5659. arXiv:cond-mat/0408348. Bibcode 2004JPCM...16.5659D. doi:10.1088/0953-8984/16/32/003.
- [20] V. A. Parsegian, *Van der Waals Forces: A Handbook for Biologists, Chemists, Engineers, and Physicists* (Cambridge Univ. Press, 2006).
- [21] Rodriguez, A. W.; Capasso, F.; Johnson, Steven G. (2011). "The Casimir effect in microstructured geometries". *Nature Photonics* **5** (4): 211–221. Bibcode 2011NaPho...5..211R. doi:10.1038/nphoton.2011.39. Review article.
- [22] B. V. Derjaguin, I. I. Abrikosova, and E. M. Lifshitz, *Quarterly Reviews, Chemical Society*, vol. 10, 295–329 (1956).
- [23] Sparnaay, M. J. (1957). "Attractive Forces between Flat Plates". *Nature* **180** (4581): 334. Bibcode 1957Natur.180..334S. doi:10.1038/180334b0.
- [24] Sparnaay, M (1958). "Measurements of attractive forces between flat plates". *Physica* **24** (6–10): 751. Bibcode 1958Phy....24..751S. doi:10.1016/S0031-8914(58)80090-7.
- [25] Movie (<http://www.drachan.com/gallery/casimir-sparnaay-werner.mov>)
- [26] Lamoreaux, S. K. (1997). "Demonstration of the Casimir Force in the 0.6 to 6  $\mu\text{m}$  Range". *Physical Review Letters* **78**: 5. Bibcode 1997PhRvL..78....5L. doi:10.1103/PhysRevLett.78.5.
- [27] Mohideen, U.; Roy, Anushree (1998). "Precision Measurement of the Casimir Force from 0.1 to 0.9  $\mu\text{m}$ ". *Physical Review Letters* **81** (21): 4549. arXiv:physics/9805038. Bibcode 1998PhRvL..81.4549M. doi:10.1103/PhysRevLett.81.4549.
- [28] Bressi, G.; Carugno, G.; Onofrio, R.; Ruoso, G. (2002). "Measurement of the Casimir Force between Parallel Metallic Surfaces". *Physical Review Letters* **88** (4): 041804. arXiv:quant-ph/0203002. Bibcode 2002PhRvL..88d1804B. doi:10.1103/PhysRevLett.88.041804. PMID 11801108.
- [29] Fulling, S. A.; Davies, P. C. W. (1976). "Radiation from a Moving Mirror in Two Dimensional Space-Time: Conformal Anomaly". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **348** (1654): 393. Bibcode 1976RSPSA.348..393F.



- doi:10.1098/rspa.1976.0045.
- [30] "First Observation of the Dynamical Casimir Effect" (<http://www.technologyreview.com/blog/arxiv/26813/>). Technology Review. .
- [31] Wilson, C. M. (2011). "Observation of the Dynamical Casimir Effect in a Superconducting Circuit". *Nature* **479**: 376-379. doi:10.1038/nature10561.
- [32] Dzyaloshinskii, I.E.; Lifshitz, E.M.; Pitaevskii, L.P. (1961). "The general theory of van der Waals forces†". *Advances in Physics* **10** (38): 165. Bibcode 1961AdPhy..10..165D. doi:10.1080/00018736100101281.
- [33] Munday, J.N.; Capasso, F.; Parsegian, V.A. (2009). "Measured long-range repulsive Casimir-Lifshitz forces". *Nature* **457** (7226): 170–3. Bibcode 2009Natur.457..170M. doi:10.1038/nature07610. PMID 19129843.
- [34] Highfield, Roger (6 August 2007). "Physicists have 'solved' mystery of levitation" (<http://www.telegraph.co.uk/news/1559579/Physicists-have-'solved'-mystery-of-levitation.html>). *The Daily Telegraph* (London). . Retrieved 28 April 2010.
- [35] Milton, Prachi; Abalo, E.K.; Parashar; Pourtolami, Nima; Brevik, Iver; Ellingsen, Simen A.. "Repulsive Casimir and Casimir-Polder Forces". *arXiv:1202.6415*, to be published in *J. Phys. A, special issue in honor of Stuart Dowker*.
- [36] Capasso, F.; Munday, J.N.; Iannuzzi, D.; Chan, H.B. (2007). "Casimir forces and quantum electrodynamic torques: physics and nanomechanics". *IEEE Journal of Selected Topics in Quantum Electronics* **13** (2): 400. doi:10.1109/JSTQE.2007.893082.
- [37] Serry, F.M.; Walliser, D.; MacLay, G.J. (1995). "The anharmonic Casimir oscillator (ACO)-the Casimir effect in a model microelectromechanical system" (<http://www.quantumfields.com/IEEEJMEMSACO.pdf>). *Journal of Microelectromechanical Systems* **4** (4): 193. doi:10.1109/84.475546. .

## Further reading

### Introductory readings

- Casimir effect description (<http://math.ucr.edu/home/baez/physics/Quantum/casimir.html>) from University of California, Riverside's version of the Usenet physics FAQ (<http://math.ucr.edu/home/baez/physics/index.html>).
- A. Lambrecht, The Casimir effect: a force from nothing (<http://physicsworld.com/cws/article/print/9747>), *Physics World*, September 2002.
- Casimir effect (<http://antwrp.gsfc.nasa.gov/apod/ap061217.html>) on Astronomy Picture of the Day

### Papers, books and lectures

- H. B. G. Casimir, and D. Polder, "The Influence of Retardation on the London-van der Waals Forces" ([http://prola.aps.org/abstract/PR/v73/i4/p360\\_1](http://prola.aps.org/abstract/PR/v73/i4/p360_1)), *Phys. Rev.* **73**, 360–372 (1948).
- Casimir, H. B. G. (1948). "On the attraction between two perfectly conducting plates" (<http://www.historyofscience.nl/search/detail.cfm?pubid=2642&view=image&startrow=1>). *Proc. Kon. Nederland. Akad. Wetensch.* **B51**: 793.
- S. K. Lamoreaux, " Demonstration of the Casimir Force in the 0.6 to 6  $\mu\text{m}$  Range (<http://link.aps.org/abstract/PRL/v78/p5>)", *Phys. Rev. Lett.* **78**, 5–8 (1997)
- M. Bordag, U. Mohideen, V.M. Mostepanenko, " New Developments in the Casimir Effect ([http://dx.doi.org/10.1016/S0370-1573\(01\)00015-1](http://dx.doi.org/10.1016/S0370-1573(01)00015-1))", *Phys. Rep.* **353**, 1–205 (2001), arXiv (<http://arxiv.org/abs/quant-ph/0106045>). (200+ page review paper.)
- Kimball A.Milton: "The Casimir effect", World Scientific, Singapore 2001,ISBN 981-02-4397-9
- Diego Dalvit, et al.: *Casimir Physics* (<http://dx.doi.org/doi:10.1007/978-3-642-20288-9>). Springer, Berlin 2011, ISBN 978-3-642-20287-2
- Bressi, G.; Carugno, G.; Onofrio, R.; Ruoso, G. (2002). "Measurement of the Casimir Force between Parallel Metallic Surfaces". *Physical Review Letters* **88** (4): 041804. arXiv:quant-ph/0203002. Bibcode 2002PhRvL..88d1804B. doi:10.1103/PhysRevLett.88.041804. PMID 11801108.
- Kenneth, O.; Klich, I.; Mann, A.; Revzen, M. (2002). "Repulsive Casimir Forces". *Physical Review Letters* **89** (3): 033001. arXiv:quant-ph/0202114. Bibcode 2002PhRvL..89c3001K. doi:10.1103/PhysRevLett.89.033001.
- J. D. Barrow, " Much ado about nothing (<http://www.gresham.ac.uk/event.asp?PageId=4&EventId=258>)", (2005) Lecture at Gresham College. (Includes discussion of French naval analogy.)

- Barrow, John D. (2000). *The book of nothing: vacuums, voids, and the latest ideas about the origins of the universe* (1st American ed.). New York: Pantheon Books. ISBN 0-09-928845-1. (Also includes discussion of French naval analogy.)
- Jonathan P. Dowling, "The Mathematics of the Casimir Effect", *Math. Mag.* **62**, 324–331 (1989).
- Patent № PCT/RU2011/000847 Author Urmatskih.

### Temperature dependence

- Measurements Recast Usual View of Elusive Force ([http://www.nist.gov/public\\_affairs/newsfromnist\\_casimir-polder.htm](http://www.nist.gov/public_affairs/newsfromnist_casimir-polder.htm)) from NIST
- V.V. Nesterenko, G. Lambiase, G. Scarpetta, Calculation of the Casimir energy at zero and finite temperature: some recent results (<http://arxiv.org/abs/hep-th/0503100>), arXiv:hep-th/0503100 v2 13 May 2005

### External links

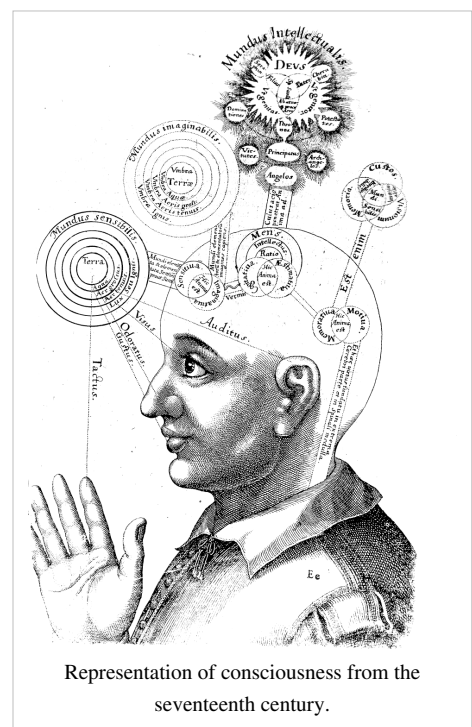
- Casimir effect article search (<http://xstructure.inr.ac.ru/x-bin/theme3.py?level=3&index1=313011>) on arxiv.org
- G. Lang, The Casimir Force (<http://www.casimir.rl.ac.uk/default.htm>) web site, 2002
- J. Babb, bibliography on the Casimir Effect (<http://www.cfa.harvard.edu/~babb/casimir-bib.html>) web site, 2009

## Consciousness

**Consciousness** is the quality or state of being aware of an external object or something within oneself.<sup>[1][2]</sup> It has been defined as: subjectivity, awareness, sentience, the ability to experience or to feel, wakefulness, having a sense of selfhood, and the executive control system of the mind.<sup>[3]</sup> Despite the difficulty in definition, many philosophers believe that there is a broadly shared underlying intuition about what consciousness is.<sup>[4]</sup> As Max Velmans and Susan Schneider wrote in *The Blackwell Companion to Consciousness*: "Anything that we are aware of at a given moment forms part of our consciousness, making conscious experience at once the most familiar and most mysterious aspect of our lives."<sup>[5]</sup>

Philosophers since the time of Descartes and Locke have struggled to comprehend the nature of consciousness and pin down its essential properties. Issues of concern in the philosophy of consciousness include whether the concept is fundamentally valid; whether consciousness can ever be explained mechanistically; whether non-human consciousness exists and if so how it can be recognized; how consciousness relates to language; whether consciousness can be understood in a way that does not require a dualistic distinction between mental and physical states or properties; and whether it may ever be possible for computers or robots to be conscious.

In recent years, consciousness has become a significant topic of research in psychology and neuroscience. The primary focus is on understanding what it means biologically and psychologically for information to be present in



Representation of consciousness from the seventeenth century.

consciousness—that is, on determining the neural and psychological correlates of consciousness. The majority of experimental studies assess consciousness by asking human subjects for a verbal report of their experiences (e.g., "tell me if you notice anything when I do this"). Issues of interest include phenomena such as subliminal perception, blindsight, denial of impairment, and altered states of consciousness produced by psychoactive drugs or spiritual or meditative techniques.

In medicine, consciousness is assessed by observing a patient's arousal and responsiveness, and can be seen as a continuum of states ranging from full alertness and comprehension, through disorientation, delirium, loss of meaningful communication, and finally loss of movement in response to painful stimuli.<sup>[6]</sup> Issues of practical concern include how the presence of consciousness can be assessed in severely ill, comatose, or anesthetized people, and how to treat conditions in which consciousness is impaired or disrupted.<sup>[7]</sup>

## In philosophy

The philosophy of mind has given rise to many stances regarding consciousness. Any attempt to impose an organization on them is bound to be somewhat arbitrary. Stuart Sutherland exemplified the difficulty in the entry he wrote for the 1989 version of the *Macmillan Dictionary of Psychology*:

**Consciousness**—The having of perceptions, thoughts, and feelings; awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means. Many fall into the trap of equating consciousness with self-consciousness—to be conscious it is only necessary to be aware of the external world. Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it has evolved. Nothing worth reading has been written on it.<sup>[8]</sup>

Most writers on the philosophy of consciousness have been concerned to defend a particular point of view, and have organized their material accordingly. For surveys, the most common approach is to follow a historical path by associating stances with the philosophers who are most strongly associated with them, for example Descartes, Locke, Kant, etc. An alternative is to organize philosophical stances according to basic issues.

## The validity of the concept

Philosophers and non-philosophers differ in their intuitions about what consciousness is.<sup>[9]</sup> While most people have a strong intuition for the existence of what they refer to as consciousness, skeptics argue that this intuition is false, either because the concept of consciousness is intrinsically incoherent, or because our intuitions about it are based in illusions. Gilbert Ryle, for example, argued that traditional understanding of consciousness depends on a Cartesian dualist outlook that improperly distinguishes between mind and body, or between mind and world. He proposed that we speak not of minds, bodies, and the world, but of individuals, or persons, acting in the world. Thus, by speaking of 'consciousness' we end up misleading ourselves by thinking that there is any sort of thing as consciousness separated from behavioral and linguistic understandings.<sup>[10]</sup> More generally, many philosophers and scientists have been unhappy about the difficulty of producing a definition that does not involve circularity or fuzziness.<sup>[8]</sup>

## Types of consciousness

Many philosophers have argued that consciousness is a unitary concept that is understood intuitively by the majority of people in spite of the difficulty in defining it.<sup>[11]</sup> Others, though, have argued that the level of disagreement about the meaning of the word indicates that it either means different things to different people, or else is an umbrella term encompassing a variety of distinct meanings with no simple element in common.<sup>[12]</sup>

Ned Block proposed a distinction between two types of consciousness that he called *phenomenal* (P-consciousness) and *access* (A-consciousness).<sup>[13]</sup> P-consciousness, according to Block, is simply raw experience: it is moving, colored forms, sounds, sensations, emotions and feelings with our bodies and responses at the center. These experiences, considered independently of any impact on behavior, are called *qualia*. A-consciousness, on the other hand, is the phenomenon whereby information in our minds is accessible for verbal report, reasoning, and the control of behavior. So, when we perceive, information about what we perceive is access conscious; when we introspect, information about our thoughts is access conscious; when we remember, information about the past is access conscious, and so on. Although some philosophers, such as Daniel Dennett, have disputed the validity of this distinction,<sup>[14]</sup> others have broadly accepted it. David Chalmers has argued that A-consciousness can in principle be understood in mechanistic terms, but that understanding P-consciousness is much more challenging: he calls this the *hard problem of consciousness*.<sup>[15]</sup>

Some philosophers believe that Block's two types of consciousness are not the end of the story. William Lycan, for example, argued in his book *Consciousness and Experience* that at least eight clearly distinct types of consciousness can be identified (organism consciousness; control consciousness; consciousness *of*; state/event consciousness; reportability; introspective consciousness; subjective consciousness; self-consciousness)—and that even this list omits several more obscure forms.<sup>[16]</sup>

## Mind–body problem

The first influential philosopher to discuss this question specifically was Descartes and the answer he gave is known as Cartesian dualism. Descartes proposed that consciousness resides within an immaterial domain he called *res cogitans* (the realm of thought), in contrast to the domain of material things which he called *res extensa* (the realm of extension).<sup>[17]</sup> He suggested that the interaction between these two domains occurs inside the brain, perhaps in a small midline structure called the pineal gland.<sup>[18]</sup>

Although it is widely accepted that Descartes explained the problem cogently, few later philosophers have been happy with his solution, and his ideas about the pineal gland have especially been ridiculed.<sup>[19]</sup> Alternative solutions, however, have been very diverse. They can be divided broadly into two categories: dualist solutions that maintain Descartes' rigid distinction between the realm of consciousness and the realm of matter but give different answers for how the two realms relate to each other; and monist solutions that maintain that there is really only one realm of being, of which consciousness and matter are both aspects. Each of these categories itself contains numerous variants. The two main types

of dualism are substance dualism (which holds that the mind is formed of a distinct type of substance not governed by the laws of physics) and property dualism (which holds that the laws of physics are universally valid but cannot

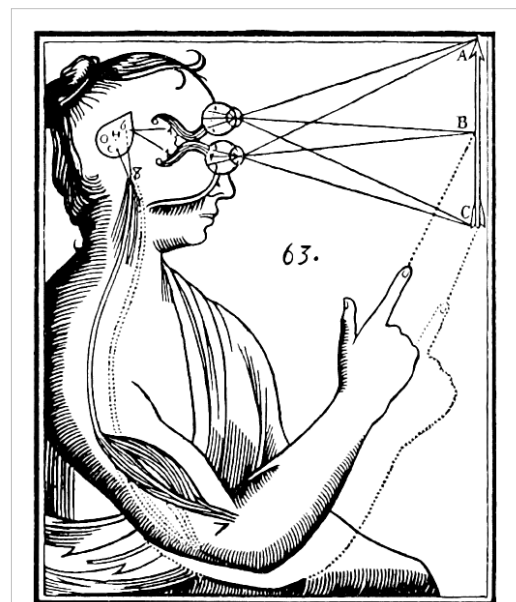


Illustration of dualism by René Descartes. Inputs are passed by the sensory organs to the pineal gland and from there to the immaterial spirit.

be used to explain the mind). The three main types of monism are physicalism (which holds that the mind consists of matter organized in a particular way), idealism (which holds that only thought truly exists and matter is merely an illusion), and neutral monism (which holds that both mind and matter are aspects of a distinct essence that is itself identical to neither of them). There are also, however, a large number of idiosyncratic theories that cannot cleanly be assigned to any of these camps.<sup>[20]</sup>

Since the dawn of Newtonian science with its vision of simple mechanical principles governing the entire universe, some philosophers have been tempted by the idea that consciousness could be explained in purely physical terms. The first influential writer to propose such an idea explicitly was Julien Offray de La Mettrie, in his book *Man a Machine* (*L'homme machine*). His arguments, however, were very abstract.<sup>[21]</sup> The most influential modern physical theories of consciousness are based on psychology and neuroscience. Theories proposed by neuroscientists such as Gerald Edelman<sup>[22]</sup> and Antonio Damasio,<sup>[23]</sup> and by philosophers such as Daniel Dennett,<sup>[24]</sup> seek to explain consciousness in terms of neural events occurring within the brain. Many other neuroscientists, such as Christof Koch,<sup>[25]</sup> have explored the neural basis of consciousness without attempting to frame all-encompassing global theories. At the same time, computer scientists working in the field of Artificial Intelligence have pursued the goal of creating digital computer programs that can simulate or embody consciousness.<sup>[26]</sup>

A few theoretical physicists have argued that classical physics is intrinsically incapable of explaining the holistic aspects of consciousness, but that quantum theory provides the missing ingredients. Several theorists have therefore proposed quantum mind (QM) theories of consciousness.<sup>[27]</sup> Notable theories falling into this category include the Holonomic brain theory of Karl Pribram and David Bohm, and the Orch-OR theory formulated by Stuart Hameroff and Roger Penrose. Some of these QM theories offer descriptions of phenomenal consciousness, as well as QM interpretations of access consciousness. None of the quantum mechanical theories has been confirmed by experiment. Recent papers by Guerreshi, G., Cia, J., Popescu, S. and Briegel, H.<sup>[28]</sup> could falsify proposals such as those of Hameroff which rely on quantum entanglement in protein. At the present time many scientists and philosophers consider the arguments for an important role of quantum phenomena to be unconvincing.<sup>[29]</sup>

Apart from the general question of the "hard problem" of consciousness, roughly speaking, the question of how mental experience arises from a physical basis,<sup>[30]</sup> a more specialized question is how to square the subjective notion that we are in control of our decisions (at least in some small measure) with the customary view of causality that subsequent events are caused by prior events. The topic of free will is the philosophical and scientific examination of this conundrum.

## Problem of other minds

Many philosophers consider experience to be the essence of consciousness, and believe that experience can only fully be known from the inside, subjectively. But if consciousness is subjective and not visible from the outside, why do the vast majority of people believe that other people are conscious, but rocks and trees are not?<sup>[31]</sup> This is called the problem of other minds.<sup>[32]</sup> It is particularly acute for people who believe in the possibility of philosophical zombies, that is, people who think it is possible in principle to have an entity that is physically indistinguishable from a human being and behaves like a human being in every way but nevertheless lacks consciousness.<sup>[33]</sup>

The most commonly given answer is that we attribute consciousness to other people because we see that they resemble us in appearance and behavior: we reason that if they look like us and act like us, they must be like us in other ways, including having experiences of the sort that we do.<sup>[34]</sup> There are, however, a variety of problems with that explanation. For one thing, it seems to violate the principle of parsimony, by postulating an invisible entity that is not necessary to explain what we observe.<sup>[34]</sup> Some philosophers, such as Daniel Dennett in an essay titled *The Unimagined Preposterousness of Zombies*, argue that people who give this explanation do not really understand what they are saying.<sup>[35]</sup> More broadly, philosophers who do not accept the possibility of zombies generally believe that consciousness is reflected in behavior (including verbal behavior), and that we attribute consciousness on the basis of behavior. A more straightforward way of saying this is that we attribute experiences to people because of what they

can *do*, including the fact that they can tell us about their experiences.<sup>[36]</sup>

### Animal consciousness

The topic of animal consciousness is beset by a number of difficulties. It poses the problem of other minds in an especially severe form, because animals, lacking the ability to express human language, cannot tell us about their experiences.<sup>[37]</sup> Also, it is difficult to reason objectively about the question, because a denial that an animal is conscious is often taken to imply that it does not feel, its life has no value, and that harming it is not morally wrong. Descartes, for example, has sometimes been blamed for mistreatment of animals due to the fact that he believed only humans have a non-physical mind.<sup>[38]</sup> Most people have a strong intuition that some animals, such as dogs, are conscious, while others, such as insects, are not; but the sources of this intuition are not obvious.<sup>[37]</sup>

Philosophers who consider subjective experience the essence of consciousness also generally believe, as a correlate, that the existence and nature of animal consciousness can never rigorously be known. Thomas Nagel spelled out this point of view in an influential essay titled *What Is it Like to Be a Bat?*. He said that an organism is conscious "if and only if there is something that it is like to be that organism — something it is like *for* the organism"; and he argued that no matter how much we know about an animal's brain and behavior, we can never really put ourselves into the mind of the animal and experience its world in the way it does itself.<sup>[39]</sup> Other thinkers, such as Douglas Hofstadter, dismiss this argument as incoherent.<sup>[40]</sup> Several psychologists and ethologists have argued for the existence of animal consciousness by describing a range of behaviors that appear to show animals holding beliefs about things they cannot directly perceive — Donald Griffin's 2001 book *Animal Minds* reviews a substantial portion of the evidence.<sup>[41]</sup>

### Artifact consciousness

The idea of an artifact made conscious is an ancient theme of mythology, appearing for example in the Greek myth of Pygmalion, who carved a statue that was magically brought to life, and in medieval Jewish stories of the Golem, a magically animated homunculus built of clay.<sup>[42]</sup> However, the possibility of actually constructing a conscious machine was probably first discussed by Ada Lovelace, in a set of notes written in 1842 about the Analytical Engine invented by Charles Babbage, a precursor (never built) to modern electronic computers. Lovelace was essentially dismissive of the idea that a machine such as the Analytical Engine could think in a humanlike way. She wrote:

It is desirable to guard against the possibility of exaggerated ideas that might arise as to the powers of the Analytical Engine. ... The Analytical Engine has no pretensions whatever to *originate* anything. It can do whatever we *know how to order it* to perform. It can *follow* analysis; but it has no power of *anticipating* any analytical relations or truths. Its province is to assist us in making *available* what we are already acquainted with.<sup>[43]</sup>

One of the most influential contributions to this question was an essay written in 1950 by pioneering computer scientist Alan Turing, titled *Computing Machinery and Intelligence*. Turing disavowed any interest in terminology, saying that even "Can machines think?" is too loaded with spurious connotations to be meaningful; but he proposed to replace all such questions with a specific operational test, which has become known as the Turing test.<sup>[44]</sup> To pass the test a computer must be able to imitate a human well enough to fool interrogators. In his essay Turing discussed a variety of possible objections, and presented a counterargument to each of them. The Turing test is commonly cited in discussions of artificial intelligence as a proposed criterion for machine consciousness; it has provoked a great deal of philosophical debate. For example, Daniel Dennett and Douglas Hofstadter argue that anything capable of passing the Turing test is necessarily conscious,<sup>[45]</sup> while David Chalmers argues that a philosophical zombie could pass the test, yet fail to be conscious.<sup>[46]</sup>

In a lively exchange over what has come to be referred to as "The Chinese room Argument", John Searle sought to refute the claim of proponents of what he calls 'Strong Artificial Intelligence (AI)' that a computer program can be conscious, though he does agree with advocates of "Weak AI" that computer programs can be formatted to

"simulate" conscious states. His own view is that consciousness has subjective, first-person causal powers by being essentially intentional due simply to the way human brains function biologically; conscious persons can perform computations, but consciousness is not inherently computational the way computer programs are. To make a Turing machine that speaks Chinese, Searle gets in a room stocked with algorithms programmed to respond to Chinese questions, i.e., Turing machines, programmed to correctly answer in Chinese questions asked in Chinese, and he finds he's able to process the inputs to outputs perfectly without having any understanding of Chinese, nor having any idea what the questions and answers could possibly mean. And, this is all a current computer program would do. If the experiment were done in English, since Searle knows English, he would be able to take questions and give answers without any algorithms for English questions, and he would be affectively aware of what was being said and the purposes it might serve: Searle passes the Turing test of answering the questions in both languages, but he's only conscious of what he's doing when he speaks English. Another way of putting the argument is to say computational computer programs can pass the Turing test for processing the syntax of a language, but that semantics cannot be reduced to syntax in the way Strong AI advocates hoped: processing semantics is conscious and intentional because we use semantics to consciously produce meaning by what we say.<sup>[47]</sup>

In the literature concerning artificial intelligence (AI), Searle's essay has been second only to Turing's in the volume of debate it has generated.<sup>[47]</sup> Searle himself was vague about what extra ingredients it would take to make a machine conscious: all he proposed was that what was needed was "causal powers" of the sort that the brain has and that computers lack. But other thinkers sympathetic to his basic argument have suggested that the necessary (though perhaps still not sufficient) extra conditions may include the ability to pass not just the verbal version of the Turing test, but the robotic version,<sup>[48]</sup> which requires grounding the robot's words in the robot's sensorimotor capacity to categorize and interact with the things in the world that its words are about, Turing-indistinguishably from a real person. Turing-scale robotics is an empirical branch of research on embodied cognition and situated cognition<sup>[49]</sup>

### **The transitivity principle**

One argument in the field of philosophy of consciousness deals with what it is that makes a mental state "conscious" in the sense of there being something it is like to experience that state. David Rosenthal posits the "transitivity principle" as a possible answer to this question. This principle holds that what makes a state conscious is the individual being aware of being in that state. This happens, on Rosenthal's account, through the use of a higher-order thought that is directed on the mental state in question.

Rosenthal cites several empirical paradigms in support of his theory. Blind-sight is one. This is a phenomenon that occurs in individuals with damage to the visual center of their brains. These individuals are often capable of relatively simple forms of visual awareness (like being able to spatially locate an x in a picture) but do not report anything concerning what it is like to experience these visual stimuli. Rosenthal claims that this can only be explained as a perception which the subject is not aware of experiencing.

Rosenthal also cites masked-priming, in which the individual is presented a priming stimulus which is quickly replaced by a masking stimulus. The individual does not report having experienced the state even though they clearly received the visual input. Again, Rosenthal claims that this can only be an instance of a visual stimulus of which the subject is not aware, and which there is therefore nothing it is like to experience.

Fred Dretske has objected to the transitivity principle on the basis that we often experience mental states that are consciously different without being aware of the conscious different. For instance, one might look at a picture of two forests. The pictures might be exactly the same except that there is one tree that is present in one picture but absent in the other. Dretske points out that what it is like to see the one forest is different from what it is like to see the other. And yet the individual looking at the pictures can easily fail to be aware that they differ at all.

## Spiritual approaches

To most philosophers, the word "consciousness" connotes the relationship between the mind and the world. To writers on spiritual or religious topics, it frequently connotes the relationship between the mind and God, or the relationship between the mind and deeper truths that are thought to be more fundamental than the physical world. Krishna consciousness, for example, is a term used to mean an intimate linkage between the mind of a worshipper and the god Krishna.<sup>[50]</sup> The mystical psychiatrist Richard Maurice Bucke distinguished between three types of consciousness: *Simple Consciousness*, awareness of the body, possessed by many animals; *Self Consciousness*, awareness of being aware, possessed only by humans; and *Cosmic Consciousness*, awareness of the life and order of the universe, possessed only by humans who are enlightened.<sup>[51]</sup> Many more examples could be given. The most thorough account of the spiritual approach may be Ken Wilber's book *The Spectrum of Consciousness*, a comparison of western and eastern ways of thinking about the mind. Wilber described consciousness as a spectrum with ordinary awareness at one end, and more profound types of awareness at higher levels.<sup>[52]</sup>

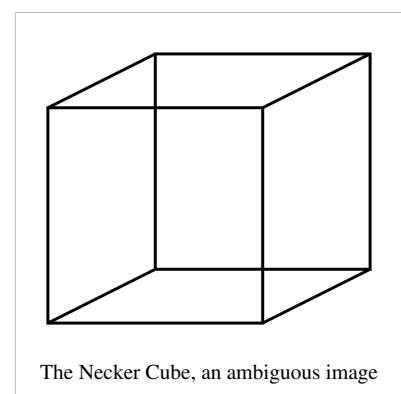
## Scientific approaches

For many decades, consciousness as a research topic was avoided by the majority of mainstream scientists, because of a general feeling that a phenomenon defined in subjective terms could not properly be studied using objective experimental methods.<sup>[53]</sup> In 1975 George Mandler published an influential psychological study which distinguished between slow, serial, and limited conscious processes and fast, parallel and extensive unconscious ones.<sup>[54]</sup> Starting in the 1980s, an expanding community of neuroscientists and psychologists have associated themselves with a field called *Consciousness Studies*, giving rise to a stream of experimental work published in books,<sup>[55]</sup> journals such as *Consciousness and Cognition*, and methodological work published in journals such as the *Journal of Consciousness Studies*, along with regular conferences organized by groups such as the Association for the Scientific Study of Consciousness.<sup>[56]</sup>

Modern scientific investigations into consciousness are based on psychological experiments (including, for example, the investigation of priming effects using subliminal stimuli), and on case studies of alterations in consciousness produced by trauma, illness, or drugs. Broadly viewed, scientific approaches are based on two core concepts. The first identifies the content of consciousness with the experiences that are reported by human subjects; the second makes use of the concept of consciousness that has been developed by neurologists and other medical professionals who deal with patients whose behavior is impaired. In either case, the ultimate goals are to develop techniques for assessing consciousness objectively in humans as well as other animals, and to understand the neural and psychological mechanisms that underlie it.<sup>[25]</sup>

## Measurement

Experimental research on consciousness presents special difficulties, due to the lack of a universally accepted operational definition. In the majority of experiments that are specifically about consciousness, the subjects are human, and the criterion that is used is verbal report: in other words, subjects are asked to describe their experiences, and their descriptions are treated as observations of the contents of consciousness.<sup>[57]</sup> For example, subjects who stare continuously at a Necker Cube usually report that they experience it "flipping" between two 3D configurations, even though the stimulus itself remains the same.<sup>[58]</sup> The objective is to understand the relationship between the conscious awareness of stimuli (as indicated by verbal report) and the effects the stimuli have on brain activity and behavior. In several paradigms,





such as the technique of response priming, the behavior of subjects is clearly influenced by stimuli for which they report no awareness.<sup>[59]</sup>

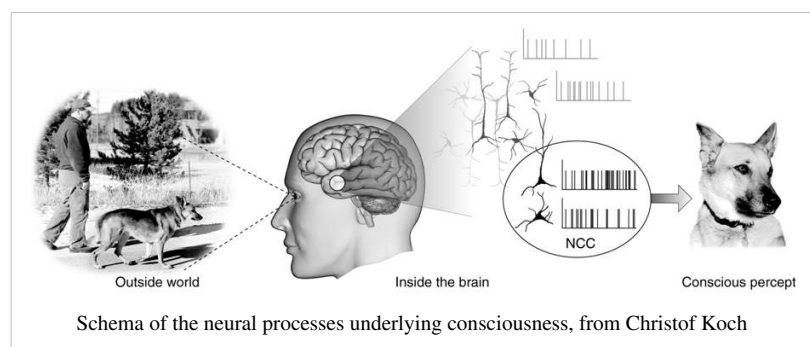
Verbal report is widely considered to be the most reliable indicator of consciousness, but it raises a number of issues.<sup>[60]</sup> For one thing, if verbal reports are treated as observations, akin to observations in other branches of science, then the possibility arises that they may contain errors—but it is difficult to make sense of the idea that subjects could be wrong about their own experiences, and even more difficult to see how such an error could be detected.<sup>[61]</sup> Daniel Dennett has argued for an approach he calls heterophenomenology, which means treating verbal reports as stories that may or may not be true, but his ideas about how to do this have not been widely adopted.<sup>[62]</sup> Another issue with verbal report as a criterion is that it restricts the field of study to humans who have language: this approach cannot be used to study consciousness in other species, pre-linguistic children, or people with types of brain damage that impair language. As a third issue, philosophers who dispute the validity of the Turing test may feel that it is possible, at least in principle, for verbal report to be dissociated from consciousness entirely: a philosophical zombie may give detailed verbal reports of awareness in the absence of any genuine awareness.<sup>[63]</sup>

Although verbal report is in practice the "gold standard" for ascribing consciousness, it is not the only possible criterion.<sup>[60]</sup> In medicine, consciousness is assessed as a combination of verbal behavior, arousal, brain activity and purposeful movement. The last three of these can be used as indicators of consciousness when verbal behavior is absent.<sup>[64]</sup> The scientific literature regarding the neural bases of arousal and purposeful movement is very extensive. Their reliability as indicators of consciousness is disputed, however, due to numerous studies showing that alert human subjects can be induced to behave purposefully in a variety of ways in spite of reporting a complete lack of awareness.<sup>[59]</sup> Studies of the neuroscience of free will have also shown that the experiences that people report when they behave purposefully sometimes do not correspond to their actual behaviors or to the patterns of electrical activity recorded from their brains.<sup>[65]</sup>

Another approach applies specifically to the study of self-awareness, that is, the ability to distinguish oneself from others. In the 1970s Gordon Gallup developed an operational test for self-awareness, known as the mirror test. The test examines whether animals are able to differentiate between seeing themselves in a mirror versus seeing other animals. The classic example involves placing a spot of coloring on the skin or fur near the individual's forehead and seeing if they attempt to remove it or at least touch the spot, thus indicating that they recognize that the individual they are seeing in the mirror is themselves.<sup>[66]</sup> Humans (older than 18 months) and other great apes, bottlenose dolphins, pigeons, and elephants have all been observed to pass this test.<sup>[67]</sup>

## Neural correlates

A major part of the scientific literature on consciousness consists of studies that examine the relationship between the experiences reported by subjects and the activity that simultaneously takes place in their brains—that is, studies of the neural correlates of consciousness. The hope is to find that activity in a particular part of the brain, or a particular pattern of global brain



activity, will be strongly predictive of conscious awareness. Several brain imaging techniques, such as EEG and fMRI, have been used for physical measures of brain activity in these studies.<sup>[68]</sup>

One idea that has drawn attention for several decades is that consciousness is associated with high-frequency (gamma band) oscillations in brain activity. This idea arose from proposals in the 1980s, by Christof von der Malsburg and Wolf Singer, that gamma oscillations could solve the so-called binding problem, by linking

information represented in different parts of the brain into a unified experience.<sup>[69]</sup> Rodolfo Llinás, for example, proposed that consciousness results from recurrent thalamo-cortical resonance where the specific thalamocortical systems (content) and the non-specific (centromedial thalamus) thalamocortical systems (context) interact in the gamma band frequency via synchronous oscillations.<sup>[70]</sup>

A number of studies have shown that activity in primary sensory areas of the brain is not sufficient to produce consciousness: it is possible for subjects to report a lack of awareness even when areas such as the primary visual cortex show clear electrical responses to a stimulus.<sup>[71]</sup> Higher brain areas are seen as more promising, especially the prefrontal cortex, which is involved in a range of higher cognitive functions collectively known as executive functions. There is substantial evidence that a "top-down" flow of neural activity (i.e., activity propagating from the frontal cortex to sensory areas) is more predictive of conscious awareness than a "bottom-up" flow of activity.<sup>[72]</sup> The prefrontal cortex is not the only candidate area, however: studies by Nikos Logothetis and his colleagues have shown, for example, that visually responsive neurons in parts of the temporal lobe reflect the visual perception in the situation when conflicting visual images are presented to different eyes (i.e., bistable percepts during binocular rivalry).<sup>[73]</sup>

In 2011 Graziano and Kastner<sup>[74]</sup> proposed the "attention schema" theory of awareness. In that theory specific cortical machinery, notably in the superior temporal sulcus and the temporo-parietal junction, is used to build the construct of awareness and attribute it to other people. The same cortical machinery is also used to attribute awareness to oneself. Damage to this cortical machinery can lead to deficits in consciousness such as hemispatial neglect. In the attention schema theory, the value of constructing the feature of awareness and attributing it to a person is to gain a useful predictive model of that person's attentional processing. Attention is a style of information processing in which a brain focuses its resources on a limited set of interrelated signals. Awareness, in this theory, is a useful, simplified schema that represents attentional state. To be aware of X is to construct a model of one's attentional focus on X.

## Defining consciousness

"The evolution of the capacity to simulate seems to have culminated in subjective consciousness. Why this should have happened is, to me, the most profound mystery facing modern biology" Richard Dawkins, *The Selfish Gene*. Since 1976, it has remained so.

In 2004, eight neuroscientists felt it was too soon for a definition. They wrote an apology in "Human Brain Function":<sup>[75]</sup>

"We have no idea how consciousness emerges from the physical activity of the brain and we do not know whether consciousness can emerge from non-biological systems, such as computers... At this point the reader will expect to find a careful and precise definition of consciousness. You will be disappointed. Consciousness has not yet become a scientific term that can be defined in this way. Currently we all use the term *consciousness* in many different and often ambiguous ways. Precise definitions of different aspects of consciousness will emerge ... but to make precise definitions at this stage is premature."

In contrast to philosophical definitions, an operational definition can be tested experimentally, and is useful for current research. A current definition for self-awareness, proposed in the 1970s by Gordon Gallup, is known as the mirror test. An operational definition proposed in 2012<sup>[76]</sup> states "consciousness is the sum of the electrical discharges occurring throughout the nervous system of a being at any given instant". What many consider consciousness may simply be the personal awareness of all the neurons delivering messages to the mind, but operational consciousness can include all neuronal activity. Extending this concept to all sentient beings, one can measure a range of consciousness based on how many and how powerfully neurons are actually firing, varying from worms to humans. One can answer the question, is someone asleep less conscious than someone thinking about a difficult problem. Although technology does not exist currently to measure this, it can be estimated by determining

oxygen consumption by the brain.

To properly understand the definition of consciousness, three principal meanings have been developed and it is critical to distinguish them. Firstly, consciousness can be defined as the waking state. This essentially means that to be conscious, one needs to be awake, aroused, alert or vigilant. The stages of consciousness can range from wakefulness, to sleep to coma even. Secondly, consciousness is defined as experience, a far more subjective approach. This notion suggests that consciousness is the content of experience from one moment to another. Consciousness is highly personal, involving a conscious subject with a limited point of view. Thirdly, consciousness can be defined as the mind. Any mental state with a propositional content is considered conscious. Thus this includes beliefs, fears, hopes, intentions, expectations and desires<sup>[77]</sup>

Christof Koch lists the following four definitions of consciousness in his latest book,<sup>[78]</sup> which can be summarized as follows:

- Consciousness is the inner mental life that we lose each night when we fall into dreamless sleep.
- Consciousness can be measured with the Glasgow Coma Scale that assesses the reactions of patients.
- An active cortico-thalamic complex is necessary for consciousness in humans, and
- Put philosophically, consciousness is what it is like to feel something.

### Biological function and evolution

Regarding the primary function of conscious processing, a recurring idea in recent theories is that phenomenal states somehow integrate neural activities and information-processing that would otherwise be independent.<sup>[79]</sup> This has been called the *integration consensus*. Another example has been proposed by Gerald Edelman called dynamic core hypothesis which puts emphasis on reentrant connections that reciprocally link areas of the brain in a massively parallel manner.<sup>[80]</sup> These theories of integrative function present solutions to two classic problems associated with consciousness: differentiation and unity. They show how our conscious experience can discriminate between infinitely different possible scenes and details (differentiation) because it integrates those details from our sensory systems, while the integrative nature of consciousness in this view easily explains how our experience can seem unified as one whole despite all of these individual parts. However, it remains unspecified which kinds of information are integrated in a conscious manner and which kinds can be integrated without consciousness. Nor is it explained what specific causal role conscious integration plays, nor why the same functionality cannot be achieved without consciousness. Obviously not all kinds of information are capable of being disseminated consciously (e.g., neural activity related to vegetative functions, reflexes, unconscious motor programs, low-level perceptual analyses, etc.) and many kinds of information can be disseminated and combined with other kinds without consciousness, as in intersensory interactions such as the ventriloquism effect.<sup>[81]</sup> Hence it remains unclear why any of it is conscious. For a review of the differences between conscious and unconscious integrations, see<sup>[81]</sup>

As noted earlier, even among writers who consider consciousness to be a well-defined thing, there is widespread dispute about which animals other than humans can be said to possess it.<sup>[82]</sup> Thus, any examination of the evolution of consciousness is faced with great difficulties. Nevertheless, some writers have argued that consciousness can be viewed from the standpoint of evolutionary biology as an adaptation in the sense of a trait that increases fitness.<sup>[83]</sup> In his paper "Evolution of consciousness," John Eccles argued that special anatomical and physical properties of the mammalian cerebral cortex gave rise to consciousness.<sup>[84]</sup> Bernard Baars proposed that once in place, this "recursive" circuitry may have provided a basis for the subsequent development of many of the functions that consciousness facilitates in higher organisms.<sup>[85]</sup> Peter Carruthers has put forth one such potential adaptive advantage gained by conscious creatures by suggesting that consciousness allows an individual to make distinctions between appearance and reality.<sup>[86]</sup> This ability would enable a creature to recognize the likelihood that their perceptions are deceiving them (e.g. that water in the distance may be a mirage) and behave accordingly, and it could also facilitate the manipulation of others by recognizing how things appear to them for both cooperative and devious ends.

Other philosophers, however, have suggested that consciousness would not be necessary for any functional advantage in evolutionary processes.<sup>[87][88]</sup> No one has given a causal explanation, they argue, of why it would not be possible for a functionally equivalent non-conscious organism (i.e., a philosophical zombie) to achieve the very same survival advantages as a conscious organism. If evolutionary processes are blind to the difference between function *F* being performed by conscious organism *O* and non-conscious organism *O\**, it is unclear what adaptive advantage consciousness could provide.<sup>[89]</sup> As a result, an exaptive explanation of consciousness has gained favor with some theorists that posit consciousness did not evolve as an adaptation but was an exaptation arising as a consequence of other developments such as increases in brain size or cortical rearrangement.

## States of consciousness

There are some states in which consciousness seems to be abolished, including sleep, coma, and death. There are also a variety of circumstances that can change the relationship between the mind and the world in less drastic ways, producing what are known as altered states of consciousness. Some altered states occur naturally; others can be produced by drugs or brain damage.<sup>[90]</sup> Altered states can be accompanied by changes in thinking, disturbances in the sense of time, feelings of loss of control, changes in emotional expression, alternations in body image and changes in meaning or significance.<sup>[91]</sup>

The two most widely accepted altered states are sleep and dreaming. Although dream sleep and non-dream sleep appear very similar to an outside observer, each is associated with a distinct pattern of brain activity, metabolic activity, and eye movement; each is also associated with a distinct pattern of experience and cognition. During ordinary non-dream sleep, people who are awakened report only vague and sketchy thoughts, and their experiences do not cohere into a continuous narrative. During dream sleep, in contrast, people who are awakened report rich and detailed experiences in which events form a continuous progression, which may however be interrupted by bizarre or fantastic intrusions. Thought processes during the dream state frequently show a high level of irrationality. Both dream and non-dream states are associated with severe disruption of memory: it usually disappears in seconds during the non-dream state, and in minutes after awakening from a dream unless actively refreshed.<sup>[92]</sup>



A Buddhist monk meditating

A variety of psychoactive drugs have notable effects on consciousness. These range from a simple dulling of awareness produced by sedatives, to increases in the intensity of sensory qualities produced by stimulants, cannabis, or most notably by the class of drugs known as psychedelics.<sup>[90]</sup> LSD, mescaline, psilocybin, and others in this group can produce major distortions of perception, including hallucinations; some users even describe their drug-induced experiences as mystical or spiritual in quality. The brain mechanisms underlying these effects are not well understood, but there is substantial evidence that alterations in the brain system that uses the chemical neurotransmitter serotonin play an essential role.<sup>[93]</sup>

There has been some research into physiological changes in yogis and people who practise various techniques of meditation. Some research with brain waves during meditation has reported differences between those corresponding to ordinary relaxation and those corresponding to meditation. It has been disputed, however, whether there is enough evidence to count these as physiologically distinct states of consciousness.<sup>[94]</sup>

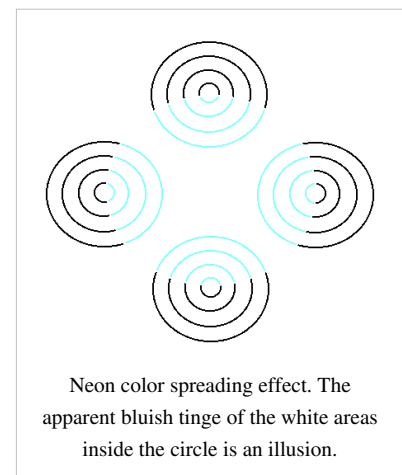
The most extensive study of the characteristics of altered states of consciousness was made by psychologist Charles Tart in the 1960s and 1970s. Tart analyzed a state of consciousness as made up of a number of component processes, including exteroception (sensing the external world); interoception (sensing the body); input-processing (seeing meaning); emotions; memory; time sense; sense of identity; evaluation and cognitive processing; motor output; and interaction with the environment.<sup>[95]</sup> Each of these, in his view, could be altered in multiple ways by drugs or other

manipulations. The components that Tart identified have not, however, been validated by empirical studies. Research in this area has not yet reached firm conclusions, but a recent questionnaire-based study identified eleven significant factors contributing to drug-induced states of consciousness: experience of unity; spiritual experience; blissful state; insightfulness; disembodiment; impaired control and cognition; anxiety; complex imagery; elementary imagery; audio-visual synesthesia; and changed meaning of percepts.<sup>[96]</sup>

## Phenomenology

Phenomenology is a method of inquiry that attempts to examine the structure of consciousness in its own right, putting aside problems regarding the relationship of consciousness to the physical world. This approach was first proposed by the philosopher Edmund Husserl, and later elaborated by other philosophers and scientists.<sup>[97]</sup> Husserl's original concept gave rise to two distinct lines of inquiry, in philosophy and psychology. In philosophy, phenomenology has largely been devoted to fundamental metaphysical questions, such as the nature of intentionality ("*aboutness*"). In psychology, phenomenology largely has meant attempting to investigate consciousness using the method of introspection, which means looking into one's own mind and reporting what one observes. This method fell into disrepute in the early twentieth century because of grave doubts about its reliability, but has been rehabilitated to some degree, especially when used in combination with techniques for examining brain activity.<sup>[98]</sup>

Introspectively, the world of conscious experience seems to have considerable structure. Immanuel Kant asserted that the world as we perceive it is organized according to a set of fundamental "intuitions", which include *object* (we perceive the world as a set of distinct things); *shape*; *quality* (color, warmth, etc.); *space* (distance, direction, and location); and *time*.<sup>[99]</sup> Some of these constructs, such as space and time, correspond to the way the world is structured by the laws of physics; for others the correspondence is not as clear. Understanding the physical basis of qualities, such as redness or pain, has been particularly challenging. David Chalmers has called this the *hard problem of consciousness*.<sup>[15]</sup> Some philosophers have argued that it is intrinsically unsolvable, because qualities ("*qualia*") are ineffable; that is, they are "raw feels", incapable of being analyzed into component processes.<sup>[100]</sup> Most psychologists and neuroscientists have not accepted these arguments — nevertheless it is clear that the relationship between a physical entity such as light and a perceptual quality such as color is extraordinarily complex and indirect, as demonstrated by a variety of optical illusions such as neon color spreading.<sup>[101]</sup>



In neuroscience, a great deal of effort has gone into investigating how the perceived world of conscious awareness is constructed inside the brain. The process is generally thought to involve two primary mechanisms: (1) hierarchical processing of sensory inputs, (2) memory. Signals arising from sensory organs are transmitted to the brain and then processed in a series of stages, which extract multiple types of information from the raw input. In the visual system, for example, sensory signals from the eyes are transmitted to the thalamus and then to the primary visual cortex; inside the cerebral cortex they are sent to areas that extract features such as three-dimensional structure, shape, color, and motion.<sup>[102]</sup> Memory comes into play in at least two ways. First, it allows sensory information to be evaluated in the context of previous experience. Second, and even more importantly, working memory allows information to be integrated over time so that it can generate a stable representation of the world—Gerald Edelman expressed this point vividly by titling one of his books about consciousness *The Remembered Present*.<sup>[103]</sup>

Despite the large amount of information available, the most important aspects of perception remain mysterious. A great deal is known about low-level signal processing in sensory systems, but the ways by which sensory systems interact with each other, with "executive" systems in the frontal cortex, and with the language system are very incompletely understood. At a deeper level, there are still basic conceptual issues that remain unresolved.<sup>[102]</sup> Many

scientists have found it difficult to reconcile the fact that information is distributed across multiple brain areas with the apparent unity of consciousness: this is one aspect of the so-called *binding problem*.<sup>[104]</sup> There are also some scientists who have expressed grave reservations about the idea that the brain forms representations of the outside world at all: influential members of this group include psychologist J. J. Gibson and roboticist Rodney Brooks, who both argued in favor of "intelligence without representation".<sup>[105]</sup>

## Medical aspects

The medical approach to consciousness is practically oriented. It derives from a need to treat people whose brain function has been impaired as a result of disease, brain damage, toxins, or drugs. In medicine, conceptual distinctions are considered useful to the degree that they can help to guide treatments. Whereas the philosophical approach to consciousness focuses on its fundamental nature and its contents, the medical approach focuses on the *amount* of consciousness a person has: in medicine, consciousness is assessed as a "level" ranging from coma and brain death at the low end, to full alertness and purposeful responsiveness at the high end.<sup>[106]</sup>

Consciousness is of concern to patients and physicians, especially neurologists and anesthesiologists. Patients may suffer from disorders of consciousness, or may need to be anesthetized for a surgical procedure. Physicians may perform consciousness-related interventions such as instructing the patient to sleep, administering general anesthesia, or inducing medical coma.<sup>[106]</sup> Also, bioethicists may be concerned with the ethical implications of consciousness in medical cases of patients such as Karen Ann Quinlan,<sup>[107]</sup> while neuroscientists may study patients with impaired consciousness in hopes of gaining information about how the brain works.<sup>[108]</sup>

## Assessment

In medicine, consciousness is examined using a set of procedures known as neuropsychological assessment.<sup>[64]</sup> There are two commonly used methods for assessing the level of consciousness of a patient: a simple procedure that requires minimal training, and a more complex procedure that requires substantial expertise. The simple procedure begins by asking whether the patient is able to move and react to physical stimuli. If so, the next question is whether the patient can respond in a meaningful way to questions and commands. If so, the patient is asked for name, current location, and current day and time. A patient who can answer all of these questions is said to be "oriented times three" (sometimes denoted "Ox3" on a medical chart), and is usually considered fully conscious.<sup>[109]</sup>

The more complex procedure is known as a neurological examination, and is usually carried out by a neurologist in a hospital setting. A formal neurological examination runs through a precisely delineated series of tests, beginning with tests for basic sensorimotor reflexes, and culminating with tests for sophisticated use of language. The outcome may be summarized using the Glasgow Coma Scale, which yields a number in the range 3—15, with a score of 3 indicating brain death (the lowest defined level of consciousness), and 15 indicating full consciousness. The Glasgow Coma Scale has three subscales, measuring the *best motor response* (ranging from "no motor response" to "obeys commands"), the *best eye response* (ranging from "no eye opening" to "eyes opening spontaneously") and the *best verbal response* (ranging from "no verbal response" to "fully oriented"). There is also a simpler pediatric version of the scale, for children too young to be able to use language.<sup>[106]</sup>

## Disorders of consciousness

Medical conditions that inhibit consciousness are considered disorders of consciousness.<sup>[110]</sup> This category generally includes minimally conscious state and persistent vegetative state, but sometimes also includes the less severe locked-in syndrome and more severe chronic coma.<sup>[110][111]</sup> Differential diagnosis of these disorders is an active area of biomedical research.<sup>[112][113][114]</sup> Finally, brain death results in an irreversible disruption of consciousness.<sup>[110]</sup> While other conditions may cause a moderate deterioration (e.g., dementia and delirium) or transient interruption (e.g., grand mal and petit mal seizures) of consciousness, they are not included in this category.

Disorder	Description
Locked-in syndrome	The patient has awareness, sleep-wake cycles, and meaningful behavior (viz., eye-movement), but is isolated due to quadriplegia and pseudobulbar palsy.
Minimally conscious state	The patient has intermittent periods of awareness and wakefulness and displays some meaningful behavior.
Persistent vegetative state	The patient has sleep-wake cycles, but lacks awareness and only displays reflexive and non-purposeful behavior.
Chronic coma	The patient lacks awareness and sleep-wake cycles and only displays reflexive behavior.
Brain death	The patient lacks awareness, sleep-wake cycles, and brain-mediated reflexive behavior.

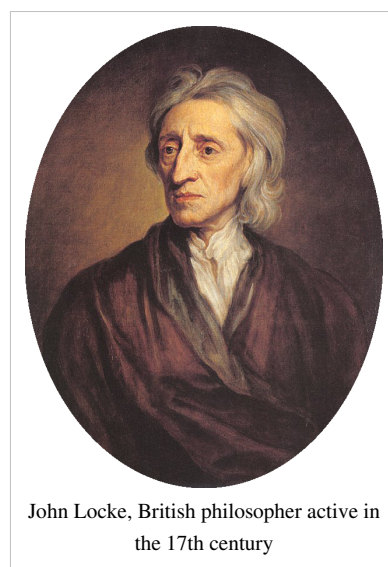
## Anosognosia

One of the most striking disorders of consciousness goes by the name anosognosia, a Greek-derived term meaning *unawareness of disease*. This is a condition in which patients are disabled in some way, most commonly as a result of a stroke, but either misunderstand the nature of the problem or deny that there is anything wrong with them.<sup>[115]</sup> The most frequently occurring form is seen in people who have experienced a stroke damaging the parietal lobe in the right hemisphere of the brain, giving rise to a syndrome known as hemispatial neglect, characterized by an inability to direct action or attention toward objects located to the right with respect to their bodies. Patients with hemispatial neglect are often paralyzed on the right side of the body, but sometimes deny being unable to move. When questioned about the obvious problem, the patient may avoid giving a direct answer, or may give an explanation that doesn't make sense. Patients with hemispatial neglect may also fail to recognize paralyzed parts of their bodies: one frequently mentioned case is of a man who repeatedly tried to throw his own paralyzed right leg out of the bed he was lying in, and when asked what he was doing, complained that somebody had put a dead leg into the bed with him. An even more striking type of anosognosia is Anton–Babinski syndrome, a rarely occurring condition in which patients become blind but claim to be able to see normally, and persist in this claim in spite of all evidence to the contrary.<sup>[116]</sup>

## Etymology and early history

The origin of the modern concept of consciousness is often attributed to John Locke's *Essay Concerning Human Understanding*, published in 1690.<sup>[117]</sup> Locke defined consciousness as "the perception of what passes in a man's own mind."<sup>[118]</sup> His essay influenced the 18th century view of consciousness, and his definition appeared in Samuel Johnson's celebrated Dictionary (1755).<sup>[119]</sup>

The earliest English language uses of "conscious" and "consciousness" date back, however, to the 1500s. The English word "conscious" originally derived from the Latin *consciūs* (*con-* "together" + *scire* "to know"), but the Latin word did not have the same meaning as our word—it meant *knowing with*, in other words *having joint or common knowledge with another*.<sup>[120]</sup> There were, however, many occurrences in Latin writings of the phrase *consciūs sibi*, which translates literally as "knowing with oneself", or in other words *sharing knowledge with oneself about something*. This phrase had the figurative meaning of *knowing that one knows*, as the modern English word



John Locke, British philosopher active in the 17th century

"conscious" does. In its earliest uses in the 1500s, the English word "conscious" retained the meaning of the Latin *consciūs*. For example Thomas Hobbes in *Leviathan* wrote: "Where two, or more men, know of one and the same

fact, they are said to be Conscious of it one to another."<sup>[121]</sup> The Latin phrase *consciūs sibi*, whose meaning was more closely related to the current concept of consciousness, was rendered in English as "conscious to oneself" or "conscious unto oneself". For example, Archbishop Ussher wrote in 1613 of "being so conscious unto myself of my great weakness".<sup>[122]</sup> Locke's definition from 1690 illustrates that a gradual shift in meaning had taken place.

A related word was *conscientia*, which primarily means moral conscience. In the literal sense, "conscientia" means knowledge-with, that is, shared knowledge. The word first appears in Latin juridical texts by writers such as Cicero.<sup>[123]</sup> Here, *conscientia* is the knowledge that a witness has of the deed of someone else.<sup>[124]</sup> René Descartes (1596–1650) is generally taken to be the first philosopher to use "conscientia" in a way that does not fit this traditional meaning.<sup>[125]</sup> Descartes used "conscientia" the way modern speakers would use "conscience." In *Search after Truth* he says "conscience or internal testimony" (*conscientia vel interno testimonio*).<sup>[126]</sup>

## Stream of consciousness

William James is usually credited with popularizing the idea that human consciousness flows like a stream, in his *Principles of Psychology* of 1890. According to James, the "stream of thought" is governed by five characteristics: "(1) Every thought tends to be part of a personal consciousness. (2) Within each personal consciousness thought is always changing. (3) Within each personal consciousness thought is sensibly continuous. (4) It always appears to deal with objects independent of itself. (5) It is interested in some parts of these objects to the exclusion of others".<sup>[127]</sup> A similar concept appears in Buddhist philosophy, expressed by the Sanskrit term *Citta-saṃtāna*, which is usually translated as mindstream or "mental continuum". In the Buddhist view, though, the "mindstream" is viewed primarily as a source of noise that distracts attention from a changeless underlying reality.<sup>[128]</sup>

In the west, the primary impact of the idea has been on literature rather than science: stream of consciousness as a narrative mode means writing in a way that attempts to portray the moment-to-moment thoughts and experiences of a character. This technique perhaps had its beginnings in the monologues of Shakespeare's plays, and reached its fullest development in the novels of James Joyce and Virginia Woolf, although it has also been used by many other noted writers.<sup>[129]</sup>

Here for example is a passage from Joyce's *Ulysses* about the thoughts of Molly Bloom:

Yes because he never did a thing like that before as ask to get his breakfast in bed with a couple of eggs since the City Arms hotel when he used to be pretending to be laid up with a sick voice doing his highness to make himself interesting for that old faggot Mrs Riordan that he thought he had a great leg of and she never left us a farthing all for masses for herself and her soul greatest miser ever was actually afraid to lay out 4d for her methylated spirit telling me all her ailments she had too much old chat in her about politics and earthquakes and the end of the world let us have a bit of fun first God help the world if all the women were her sort down on bathingsuits and lownecks of course nobody wanted her to wear them I suppose she was pious because no man would look at her twice I hope Ill never be like her a wonder she didnt want us to cover our faces but she was a welleducated woman certainly and her gabby talk about Mr Riordan here and Mr Riordan there I suppose he was glad to get shut of her.<sup>[130]</sup>



## References

- [1] "consciousness" (<http://www.merriam-webster.com/dictionary/consciousness>). Merriam-Webster. . Retrieved June 4, 2012.
- [2] Robert van Gulick (2004). "Consciousness" (<http://plato.stanford.edu/entries/consciousness/>). Stanford Encyclopedia of Philosophy. .
- [3] Farthing G (1992). *The Psychology of Consciousness*. Prentice Hall. ISBN 978-0-13-728668-3.
- [4] John Searle (2005). "Consciousness". In Honderich T. *The Oxford companion to philosophy*. Oxford University Press. ISBN 978-0-19-926479-7.
- [5] Susan Schneider and Max Velmans (2008). "Introduction". In Max Velmans, Susan Schneider. *The Blackwell Companion to Consciousness*. Wiley. ISBN 978-0-470-75145-9.
- [6] Güven Güzeldere (1997). Ned Block, Owen Flanagan, Güven Güzeldere. ed. *The Nature of Consciousness: Philosophical debates*. Cambridge, MA: MIT Press. pp. 1–67.
- [7] J. J. Fins, N. D. Schiff, and K. M. Foley (2007). "Late recovery from the minimally conscious state: ethical and policy implications". *Neurology* **68** (4): 304–307. doi:10.1212/01.wnl.0000252376.43779.96. PMID 17242341.
- [8] Stuart Sutherland (1989). "Consciousness". *Macmillan Dictionary of Psychology*. Macmillan. ISBN 978-0-333-38829-7.
- [9] Justin Sytsma and Edouard Machery (2010). "Two conceptions of subjective experience". *Philosophical Studies* **151**: 299–327. doi:10.1007/s11098-009-9439-x.
- [10] Gilbert Ryle (1949). *The Concept of Mind*. University of Chicago Press. pp. 156–163. ISBN 978-0-226-73296-1.
- [11] Michael V. Antony (2001). "Is consciousness ambiguous?". *Journal of Consciousness Studies* **8**: 19–44.
- [12] Max Velmans (2009). "How to define consciousness—and how not to define consciousness". *Journal of Consciousness Studies* **16**: 139–156.
- [13] Ned Block (1998). "On a confusion about a function of consciousness" (<http://cogprints.org/231/1/199712004.html>). In N. Block, O. Flanagan, G. Guzeldere. *The Nature of Consciousness: Philosophical Debates*. MIT Press. pp. 375–415. ISBN 978-0-262-52210-6. .
- [14] Daniel Dennett (2004). *Consciousness Explained*. Penguin. pp. 375. ISBN 0-713-99037-6.
- [15] David Chalmers (1995). "Facing up to the problem of consciousness" (<http://www.imprint.co.uk/chalmers.html>). *Journal of Consciousness Studies* **2**: 200–219. .
- [16] William Lycan (1996). *Consciousness and Experience*. MIT Press. pp. 1–4. ISBN 978-0-262-12197-2.
- [17] Dy, Jr., Manuel B. (2001). *Philosophy of Man: selected readings*. Goodwill Trading Co.. p. 97. ISBN 971-12-0245-X.
- [18] "Descartes and the Pineal Gland" (<http://plato.stanford.edu/entries/pineal-gland/>). Stanford University. November 5, 2008. . Retrieved 2010-08-22.
- [19] Gert-Jan Lokhorst. "Descartes and the pineal gland" (<http://plato.stanford.edu/archives/sum2011/entries/pineal-gland/>). In Edward N. Zalta. Stanford Encyclopedia of Philosophy (Summer 2011 Edition). .
- [20] William Jaworski (2011). *Philosophy of Mind: A Comprehensive Introduction*. John Wiley and Sons. pp. 5–11. ISBN 978-1-4443-3367-1.
- [21] Julien Offray de La Mettrie (1996). Ann Thomson. ed. *Machine man and other writings*. Cambridge University Press. ISBN 978-0-521-47849-6.
- [22] Gerald Edelman (1993). *Bright Air, Brilliant Fire: On the Matter of the Mind*. Basic Books. ISBN 978-0-465-00764-6.
- [23] Antonio Damasio (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Press. ISBN 978-0-15-601075-7.
- [24] Daniel Dennett (1991). *Consciousness Explained*. Boston: Little & Company. ISBN 978-0-316-18066-5.
- [25] Christof Koch (2004). *The Quest for Consciousness*. Englewood CO: Roberts & Company. ISBN 978-0-9747077-0-9.
- [26] Stuart J. Russell, Peter Norvig (2010). "Ch. 26: Philosophical foundations". *Artificial Intelligence: A Modern Approach*. Prentice Hall. ISBN 978-0-13-604259-4.
- [27] "Quantum Approaches to Consciousness" (<http://plato.stanford.edu/entries/qt-consciousness/>). Stanford University. December 25, 2011. .
- [28] Cai, J.; Popescu, S.; Briegel, H.. "Persistent dynamic entanglement from classical motion: How bio-molecular machines can generate non-trivial quantum states". *Physical Review E* **82**: 021921. doi:10.1103/PhysRevE.82.021921.
- [29] John Searle (1997). *The Mystery of Consciousness*. The New York Review of Books. pp. 53–88. ISBN 978-0-940322-06-6.
- [30] For a discussion see Rocco J. Gennaro (2011). "\$4.4 The hard problem of consciousness" (<http://books.google.com/books?id=t-XgKMgzwk4C&pg=PA75&lpg=PA75>). *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. MIT Press. p. 75. ISBN 0262016605. .
- [31] Knobe J (2008). "Can a Robot, an Insect or God Be Aware?" (<http://www.scientificamerican.com/article.cfm?id=can-a-robot-an-insect-or>). *Scientific American: Mind*. .
- [32] Alec Hyslop (1995). *Other Minds*. Springer. pp. 5–14. ISBN 978-0-7923-3245-9.
- [33] Robert Kirk. "Zombies" (<http://plato.stanford.edu/archives/sum2009/entries/zombies>). In Edward N. Zalta. Stanford Encyclopedia of Philosophy (Summer 2009 Edition). .
- [34] Alec Hyslop (1995). "The analogical inference to other minds". *Other Minds*. Springer. pp. 41–70. ISBN 978-0-7923-3245-9.
- [35] Daniel Dennett (1995). "The unimagined preposterousness of zombies". *Journal of Consciousness Studies* **2**: 322–325.
- [36] Stevan Harnad (1995). "Why and how we are not zombies". *Journal of Consciousness Studies* **1**: 164–167.
- [37] Colin Allen. "Animal consciousness" (<http://plato.stanford.edu/archives/sum2011/entries/consciousness-animal/>). In Edward N. Zalta. Stanford Encyclopedia of Philosophy (Summer 2011 Edition). .

- [38] Peter Carruthers (1999). "Sympathy and subjectivity". *Australasian Journal of Philosophy* **77**: 465–482.
- [39] Thomas Nagel (1991). "Ch. 12 What is it like to be a bat?". *Mortal Questions*. Cambridge University Press. ISBN 978-0-521-40676-5.
- [40] Douglas Hofstadter (1981). "Reflections on *What Is It Like to Be a Bat?*". In Douglas Hofstadter and Daniel Dennett. *The Mind's I*. Basic Books. pp. 403–414. ISBN 046504624.
- [41] Donald Griffin (2001). *Animal Minds: Beyond Cognition to Consciousness*. University of Chicago Press. ISBN 978-0-226-30865-4.
- [42] Moshe Idel (1990). *Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid*. SUNY Press. ISBN 978-0-7914-0160-6.  
Note: In many stories the Golem was mindless, but some gave it emotions or thoughts.
- [43] Ada Lovelace. "Sketch of The Analytical Engine, Note G" (<http://www.fourmilab.ch/babbage/sketch.html>). .
- [44] Stuart Shieber (2004). *The Turing Test : Verbal Behavior as the Hallmark of Intelligence*. MIT Press. ISBN 978-0-262-69293-9.
- [45] Daniel Dennett and Douglas Hofstadter (1985). *The Mind's I*. Basic Books. ISBN 978-0-553-34584-1.
- [46] David Chalmers (1997). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press. ISBN 019511789.
- [47] "The Chinese Room Argument" (<http://plato.stanford.edu/entries/chinese-room/>). .
- [48] Graham Oppy and David Dowe (2011). "The Turing test" (<http://plato.stanford.edu/archives/spr2011/entries/turing-test/>). Stanford Encyclopedia of Philosophy (Spring 2011 Edition). .
- [49] Margaret Wilson (2002). "Six views of embodied cognition" (<http://philosophy.wisc.edu/shapiro/PHIL951/951articles/wilson.htm>). *Psychonomic Bulletin & Review* **9**: 625–636. doi:10.3758/BF03196322. .
- [50] Lynne Gibson (2002). *Modern World Religions: Hinduism*. Heinemann Educational Publishers. pp. 2–4. ISBN 0-435-33619-3.
- [51] Richard Maurice Bucke (1905). *Cosmic Consciousness: A Study in the Evolution of the Human Mind* (<http://books.google.com/books?id=gxRWAAAAMAAJ>). Innes & Sons. pp. 1–2. .
- [52] Ken Wilber (2002). *The Spectrum of Consciousness*. Motilal Banarsidass Publ.. pp. 3–16. ISBN 978-81-208-1848-4.
- [53] Horst Hendriks-Jansen (1996). *Catching ourselves in the act: situated activity, interactive emergence, evolution, and human thought*. Massachusetts Institute of Technology. p. 114. ISBN 0-262-08246-2.
- [54] Mandler, G. Consciousness: Respectable, useful, and probably necessary. In R.Solso (Ed.) *Information processing and cognition*: NJ: LEA.
- [55] Mandler, G. *Consciousness recovered: Psychological functions and origins of thought*. Philadelphia: John Benjamins. 2002
- [56] Stuart Hameroff, Alfred Kaszniak, David Chalmers (1999). "Preface". *Toward a Science of Consciousness III: The Third Tucson Discussions and Debates*. MIT Press. pp. xix–xx. ISBN 978-0-262-58181-3.
- [57] Bernard Baars (1993). *A Cognitive Theory of Consciousness*. Cambridge University Press. pp. 15–18. ISBN 978-0-521-42743-2.
- [58] Paul Rooks and Jane Wilson (2000). *Perception: Theory, Development, and Organization*. Psychology Press. pp. 25–26. ISBN 978-0-415-19094-7.
- [59] Thomas Schmidt and Dirk Vorberg (2006). "Criteria for unconscious cognition: Three types of dissociation". *Perception and Psychophysics* **68**: 489–504.
- [60] Arnaud Destrebecqz and Philippe Peigneux (2006). "Methods for studying unconscious learning". In Steven Laureys. *The Boundaries of Consciousness: Neurobiology and Neuropathology*. Elsevier. pp. 69–80. ISBN 978-0-444-52876-6.
- [61] Daniel Dennett (1992). "Quining qualia" (<http://cogprints.org/254/>). In A. Marcel and E. Bisiach. *Consciousness in Modern Science*. Oxford University Press. ISBN 978-0-19-852237-9. . Retrieved 2011-10-31.
- [62] Daniel Dennett (2003). "Who's on first? Heterophenomenology explained". *Journal of Consciousness Studies* **10**: 19–30.
- [63] David Chalmers (1996). "Ch. 3: Can consciousness be reductively explained?". *The Conscious Mind*. Oxford University Press. ISBN 978-0-19-511789-9.
- [64] J. T. Giacino, C. M. Smart (2007). "Recent advances in behavioral assessment of individuals with disorders of consciousness". *Current Opinion in Neurology* **20**: 614–619. doi:10.1097/WCO.0b013e3282f189ef. PMID 17992078.
- [65] Patrick Haggard (2008). "Human volition: towards a neuroscience of will". *Nature Reviews Neuroscience* **9** (12): 934–946. doi:10.1038/nrn2497. PMID 19020512.
- [66] Gordon Gallup (1970). "Chimpanzees: Self recognition". *Science* **167** (3914): 86–87. doi:10.1126/science.167.3914.86. PMID 4982211.
- [67] David Edelman and Anil Seth (2009). "Animal consciousness: a synthetic approach". *Trends in Neurosciences* **32**: 476–484. doi:10.1016/j.tins.2009.05.008. PMID 19716185.
- [68] Christof Koch (2004). *The Quest for Consciousness*. Englewood CO: Roberts & Company. pp. 16–19. ISBN 978-0-9747077-0-9.
- [69] Wolf Singer. "Binding by synchrony" ([http://www.scholarpedia.org/article/Binding\\_by\\_synchrony](http://www.scholarpedia.org/article/Binding_by_synchrony)). Scholarpedia. . Retrieved 2011-10-26.
- [70] Rodolfo Llinás (2002). *I of the Vortex. From Neurons to Self*. MIT Press. ISBN 978-0-262-62163-2.
- [71] Koch, *The Quest for Consciousness*, pp. 105–116
- [72] Francis Crick and Christof Koch (2003). "A framework for consciousness" (<http://papers.klab.caltech.edu/29/1/438.pdf>) (PDF). *Nature Neuroscience* **6**: 119–126. doi:10.1038/nn0203-119. PMID 12555104. .
- [73] Koch, *The Quest for Consciousness*, pp. 269–286
- [74] Graziano, M.S.A. and Kastner, S (2011). "Human consciousness and its relationship to social neuroscience: A novel hypothesis". *Cog. Neurosci* **2**: 98–113.
- [75] *Human Brain Function*, by Richard Frackowiak and 7 other neuroscientists, page 269 in chapter 16 "The Neural Correlates of Consciousness" (consisting of 32 pages), published 2004
- [76] What Consciousness is (<http://chemistry.beloit.edu/Ordman/classes/cls/idst125sp11/1rocdef.htm>) Roc Ordman's Consciousness class, Beloit College, 2012,

- [77] Zeman, Adam (July 2001). "Consciousness" (<http://simplelink.library.utoronto.ca/url.cfm/329042>). *Journal of Neurology* **124** (7): 1263–1289. . Retrieved 08/12/2012.
- [78] Consciousness, Confessions of a Romantic Reductionist, Christof Koch, 2012, MIT Press
- [79] Bernard Baars. "The conscious access hypothesis: Origins and recent evidence". *Trends in Cognitive Sciences* **6**: 47–52. doi:10.1016/S1364-6613(00)01819-2. PMID 11849615.
- [80] Seth, Anil; Eugene Izhikevich, George Reeke, Gerald Edelman (2006). "Theories and measures of consciousness: An extended framework". *Proceedings of the National Academy of Sciences* **103** (28). doi:10.1073/pnas.0604347103.
- [81] Ezequiel Morsella (2005). "The function of phenomenal states: Supramodular Interaction Theory". *Psychological Review* **112** (4): 1000–1021. doi:10.1037/0033-295X.112.4.1000. PMID 16262477.
- [82] S. Budioansky (1998). *If a Lion Could Talk: Animal Intelligence and the Evolution of Consciousness*. The Free Press. ISBN 978-0-684-83710-9.
- [83] S. Nichols, T. Grantham (2000). "Adaptive Complexity and Phenomenal Consciousness". *Philosophy of Science* **67**: 648–670. doi:10.1086/392859.
- [84] John Eccles (1992). "Evolution of consciousness". *Proc. Natl. Acad. Sci. U.S.A.* **89**: 7320–7324. doi:10.1073/pnas.89.16.7320. PMC 49701. PMID 1502142.
- [85] Bernard Baars (1993). *A Cognitive Theory of Consciousness*. Cambridge University Press. ISBN 978-0-521-42743-2.
- [86] Carruthers, Peter (2004). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge, UK: Cambridge University Press.
- [87] Owen Flanagan and T. W. Polger (1995). "Zombies and the function of consciousness". *Journal of Consciousness Studies* **2**: 313–321.
- [88] Rosenthal, David (2008). "Consciousness and its function". *Neuropsychologia* **46**: 829–840. doi:10.1016/j.neuropsychologia.2007.11.012.
- [89] Stevan Harnad (2002). "Turing indistinguishability and the Blind Watchmaker" (<http://cogprints.org/1615>). In J. H. Fetzer. *Consciousness Evolving*. John Benjamins. . Retrieved 2011-10-26.
- [90] Dieter Vaitl *et al.* (2005). "Psychobiology of altered states of consciousness". *Psychological Bulletin* **131**: 98–127. doi:10.1037/0033-2909.131.1.98. PMID 15631555.
- [91] Schacter, Daniel; Gilbert, Daniel; Wegner, Daniel (2011). *Psychology 2nd Ed.*. 41 Madison Avenue, New York, NY 10010: Worth Publishers. pp. 190. ISBN 1-4292-3719-8.
- [92] J. Allan Hobson, Edward F. Pace-Schott, and Robert Stickgold (2003). "Dreaming and the brain: Toward a cognitive neuroscience of conscious states". In Edward F. Pace-Schott, Mark Solms, Mark Blagrove, Stevan Harnad. *Sleep and Dreaming: Scientific Advances and Reconsiderations*. Cambridge University Press. ISBN 978-0-521-00869-3.
- [93] Michael Lyvers (2003). "The neurochemistry of psychedelic experiences" ([http://epublications.bond.edu.au/hss\\_pubs/10](http://epublications.bond.edu.au/hss_pubs/10)) (PDF). ePublications@bond. .
- [94] M. Murphy, S. Donovan, and E. Taylor (1997). *The Physical and Psychological Effects of Meditation: A Review of Contemporary Research With a Comprehensive Bibliography, 1931-1996*. Institute of Noetic Sciences.
- [95] Charles Tart (2001). "Ch. 2: The components of consciousness" (<http://www.psychedelic-library.org/soc2.htm>). *States of Consciousness*. IUniverse.com. ISBN 978-0-595-15196-7. . Retrieved 2011-10-05.
- [96] E. Studerus, A. Gamma, and F. X. Vollenweider (2010). "Psychometric evaluation of the altered states of consciousness rating scale (OAV)". *PLoS ONE* **8**. doi:10.1371/journal.pone.0012412. PMC 2930851. PMID 20824211.
- [97] Robert Sokolowski (2000). *Introduction to Phenomenology*. Cambridge University Press. pp. 211–227. ISBN 978-0-521-66792-0.
- [98] K. Anders Ericsson (2003). "Valid and non-reactive verbalization of thoughts during performance of tasks: towards a solution to the central problems of introspection as a source of scientific evidence". In Anthony Jack, Andreas Roepstorff. *Trusting the Subject?: The Use of Introspective Evidence in Cognitive Science, Volume 1*. Imprint Academic. pp. 1–18. ISBN 978-0-907845-56-0.
- [99] Andrew Brook. "Kant's view of the mind and consciousness of self" (<http://plato.stanford.edu/entries/kant-mind/#3.2.1>). Stanford Encyclopedia of Philosophy. . Note: translating Kant's terminology into English is often difficult.
- [100] Joseph Levine (1998). "On leaving out what it's like". In N. Block, O. Flanagan, G. Guzeldere. *The Nature of Consciousness: Philosophical Debates*. MIT Press. ISBN 978-0-262-52210-6.
- [101] Steven K. Shevell (2003). "Color appearance". In Steven K. Shevell. *The Science of Color*. Elsevier. pp. 149–190. ISBN 978-0-444-51251-2.
- [102] M. R. Bennett and Peter Michael Stephan Hacker (2003). *Philosophical Foundations of Neuroscience*. Wiley-Blackwell. pp. 121–147. ISBN 978-1-4051-0838-6.
- [103] Gerald Edelman (1989). *The Remembered Present: A Biological Theory of Consciousness*. Basic Books. pp. 109–118. ISBN 978-0-465-06910-1.
- [104] Koch, *The Quest for Consciousness*, pp. 167–170
- [105] Rodney Brooks (1991). "Intelligence without representation". *Artificial Intelligence* **47**: 139–159. doi:10.1016/0004-3702(91)90053-M.
- [106] Hal Blumenfeld (2009). "The neurological examination of consciousness". In Steven Laureys, Giulio Tononi. *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*. Academic Press. ISBN 978-0-12-374168-4.
- [107] Kinney HC, Korein J, Panigrahy A, Dikkes P, Goode R (26 May 1994). "Neuropathological findings in the brain of Karen Ann Quinlan -- the role of the thalamus in the persistent vegetative state". *N Engl J Med* **330** (21): 1469–1475. doi:10.1056/NEJM199405263302101. PMID 8164698.
- [108] Koch, *The Quest for Consciousness*, pp. 216–226


- [109] V. Mark Durand and David H. Barlow (2009). *Essentials of Abnormal Psychology*. Cengage Learning. pp. 74–75. ISBN 978-0-495-59982-1. Note: A patient who can additionally describe the current situation may be referred to as "oriented times four".
- [110] Bernat JL (8 Apr 2006). "Chronic disorders of consciousness". *Lancet* **367** (9517): 1181–1192. doi:10.1016/S0140-6736(06)68508-5. PMID 16616561.
- [111] Bernat JL (20 Jul 2010). "The natural history of chronic disorders of consciousness". *Neurol* **75** (3): 206–207. doi:10.1212/WNL.0b013e3181e8e960. PMID 20554939.
- [112] Coleman MR, Davis MH, Rodd JM, Robson T, Ali A, Owen AM, Pickard JD (September 2009). "Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness". *Brain* **132** (9): 2541–2552. doi:10.1093/brain/awp183. PMID 19710182.
- [113] Monti MM, Vanhaudenhuyse A, Coleman MR, Boly M, Pickard JD, Tshibanda L, Owen AM, Laureys S (18 Feb 2010). "Willful modulation of brain activity in disorders of consciousness". *N Engl J Med* **362** (7): 579–589. doi:10.1056/NEJMoa0905370. PMID 20130250.
- [114] Seel RT, Sherer M, Whyte J, Katz DI, Giacino JT, Rosenbaum AM, Hammond FM, Kalmar K, Pape TL, et al. (December 2010). "Assessment scales for disorders of consciousness: evidence-based recommendations for clinical practice and research". *Arch Phys Med Rehabil* **91** (12): 1795–1813. doi:10.1016/j.apmr.2011.01.002. PMID 21112421.
- [115] George P. Prigatano and Daniel Schacter (1991). "Introduction". In George Prigatano, Daniel Schacter. *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issues*. Oxford University Press. pp. 3–16. ISBN 0-19-505941-7.
- [116] Kenneth M. Heilman (1991). "Anosognosia: possible neuropsychological mechanisms". In George Prigatano, Daniel Schacter. *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issues*. Oxford University Press. pp. 53–62. ISBN 0-19-505941-7.
- [117] Locke, John. "An Essay Concerning Human Understanding (Chapter XXVII)" (<http://ebooks.adelaide.edu.au/l/locke/john/181u/B2.27.html>). Australia: University of Adelaide. . Retrieved August 20, 2010.
- [118] "Science & Technology: consciousness" (<http://www.britannica.com/EBchecked/topic/133274/consciousness>). Encyclopædia Britannica. . Retrieved August 20, 2010.
- [119] Samuel Johnson (1756). *A Dictionary of the English Language* (<http://books.google.com/books?id=fcVEAAAaAAJ>). Knapton. .
- [120] C. S. Lewis (1990). "Ch. 8: Conscience and conscious". *Studies in words*. Cambridge University Press. ISBN 978-0-521-39831-2.
- [121] Thomas Hobbes (1904). *Leviathan: or, The Matter, Forme & Power of a Commonwealth, Ecclesiasticall and Civill* (<http://books.google.com/books?id=2oc6AAAAMAAJ>). University Press. p. 39. .
- [122] James Ussher, Charles Richard Elrington (1613). *The whole works, Volume 2*. Hodges and Smith. p. 417.
- [123] James Hastings and John A. Selbie (2003). *Encyclopedia of Religion and Ethics Part 7*. Kessinger Publishing. p. 41. ISBN 0-7661-3677-9.
- [124] G. Melenaar. *Mnemosyne, Fourth Series*. **22**. Brill. pp. 170–180.
- [125] Boris Hennig (2007). "Cartesian Conscientia". *British Journal for the History of Philosophy* **15**: 455–484.
- [126] Sara Heinämaa, Vili Lähteenmäki, Pauliina Remes (eds.) (2007). *Consciousness: from perception to reflection in the history of philosophy*. Springer. pp. 205–206. ISBN 978-1-4020-6081-6.
- [127] William James (1890). *The Principles of Psychology, Volume 1*. H. Holt. p. 225.
- [128] Dzogchen Rinpoche (2007). "Taming the mindstream". In Doris Wolter. *Losing the Clouds, Gaining the Sky: Buddhism and the Natural Mind*. Wisdom Publications. pp. 81–92. ISBN 978-0-86171-359-2.
- [129] Robert Humphrey (1954). *Stream of Consciousness in the Modern Novel*. University of California Press. pp. 23–49. ISBN 978-0-520-00585-3.
- [130] James Joyce (1990). *Ulysses*. Bompacrazy.com. p. 620.

## External links

- *Anthropology of Consciousness* (<http://www.wiley.com/bw/journal.asp?ref=1053-4202>)
- *Consciousness and Cognition* ([http://www.elsevier.com/wps/find/journaldescription.cws\\_home/622810/description#description](http://www.elsevier.com/wps/find/journaldescription.cws_home/622810/description#description))
- *Consciousness & Emotion* (<http://www.ingentaconnect.com/content/jbp/ce>)
- Consciousness Studies ([http://www.dmoz.org/Society/Philosophy/Philosophy\\_of\\_Mind/Consciousness\\_Studies/](http://www.dmoz.org/Society/Philosophy/Philosophy_of_Mind/Consciousness_Studies/)) at the Open Directory Project
- *Journal of Consciousness Studies* (<http://www.imprint.co.uk/jcs.html>)
- *Psyche* ([http://www.theassc.org/journal\\_psyche](http://www.theassc.org/journal_psyche)) (ASSC)
- Quantum Mind (<http://www.quantum-mind.co.uk>)
- Evolution and Function of Consciousness (<http://turingc.blogspot.ca>) Alan Turing Year

# Stuart Hameroff

---

<b>Stuart Hameroff</b>	
	
<b>Born</b>	16 July 1947 Buffalo, New York, USA
<b>Residence</b>	Tucson, Arizona, USA
<b>Citizenship</b>	United States
<b>Nationality</b>	American
<b>Fields</b>	Anesthesiologist
<b>Institutions</b>	University of Arizona
<b>Alma mater</b>	University of Pittsburgh (B.S.) Hahnemann University Hospital (M.D.)
<b>Known for</b>	Consciousness studies
<b>Influences</b>	Roger Penrose

**Stuart Hameroff** (born on July 16, 1947) is an anesthesiologist and professor at the University of Arizona known for his studies of consciousness.

## Career

Hameroff received his BS degree from the University of Pittsburgh and his MD degree from Hahnemann University Hospital, where he studied before it became part of the Drexel University College of Medicine. He took an internship at the Tucson Medical Center in 1973. From 1975 onwards, he has spent the whole of his career at the University of Arizona, becoming professor in the Department of Anesthesiology and Psychology and associate director for the Center for Consciousness Studies, both in 1999, and finally Emeritus professor for Anesthesiology and Psychology in 2003.

---

## Hypotheses

At the very beginning of Dr. Hameroff's career, while he was at Hahnemann, cancer-related research work piqued his interest in the part played by microtubules in cell division, and led him to speculate that they were controlled by some form of computing. It also suggested to him that part of the solution of the problem of consciousness might lie in understanding the operations of microtubules in brain cells, operations at the molecular and supramolecular level<sup>[1]</sup>.

The operations of microtubules are remarkably complex and their role pervasive in cellular operations; these facts led to the speculation that computation sufficient for consciousness might somehow be occurring there. These ideas are discussed in Hameroff's first book *Ultimate Computing* (1987)<sup>[2]</sup>. The main substance of this book dealt with the scope for information processing in biological tissue and especially in microtubules and other parts of the cytoskeleton. Hameroff argued that these subneuronal cytoskeleton components could be the basic units of processing rather than the neurons themselves. The book was primarily concerned with information processing, with consciousness secondary at this stage.

Separately from Hameroff, Roger Penrose had published his first book on consciousness, *The Emperor's New Mind*<sup>[3]</sup>. On the basis of Godel's incompleteness theorems, he argued that the brain could perform functions that no computer or system of algorithms could. From this it could follow that consciousness itself might be fundamentally non-algorithmic, and incapable of being modelled as a classical Turing machine type of computer. By contrast, the idea that it could be explained mechanistically was prevalent in the field of Artificial Intelligence at that time.

Penrose saw the principles of quantum theory as providing an alternative process through which consciousness could arise. He further argued that this non-algorithmic process in the brain required a new form of the quantum wave reduction, later given the name objective reduction (OR), which could link the brain to the fundamental spacetime geometry. At this stage, he had no precise ideas as to how such a quantum process might be instantiated in the brain. Moreover, Penrose's ideas were widely criticized by neuroscientists, logicians and philosophers, notably Grush and Churchland (Grush and Churchland, 1995)<sup>[4]</sup>.

Hameroff was inspired by Penrose's book to contact Penrose regarding his own theories about the mechanism of anesthesia, and how it specifically targets consciousness via action on neural microtubules. The two met in 1992, and Hameroff suggested that the microtubules were a good candidate site for a quantum mechanism in the brain. Penrose was interested in the mathematical features of the microtubule lattice, and over the next two years the two collaborated in formulating the orchestrated objective reduction (Orch-OR) model of consciousness<sup>[1]</sup>. Following this collaboration, Penrose published his second consciousness book, *Shadows of the Mind*<sup>[5]</sup>.

This more developed version of their ideas was also widely attacked, and notably by the physicist Max Tegmark, who calculated that quantum states in microtubules would survive for only  $10^{-13}$  seconds, too brief to be of any significance for neural processes (Tegmark, 2000)<sup>[6]</sup>. Hameroff and the physicists Scott Hagan and Jack Tuszynski (Hagan, Hameroff & Tuszynski, 2002)<sup>[7]</sup> replied to Tegmark arguing that microtubules could be shielded against the environment of the brain. To date, there is no experimental confirmation of these proposed methods of shielding, but Hameroff has proposed tests that could falsify the theory<sup>[8]</sup>.

Over the years since 1994, Hameroff has been active in promoting the Orch-OR model of consciousness through his web site<sup>[1]</sup>, conferences and lectures. He was the lead organizer of the first Tucson consciousness meeting in 1994 that brought together approximately 300 people interested in consciousness studies (e.g., David Chalmers, Christof Koch, Bernard Baars, Roger Penrose, Benjamin Libet). This conference is widely regarded as a landmark event within the field of consciousness studies, and by bringing researchers from various disciplines together led to various useful synergies, resulting indirectly, for instance, in the formation of the Association for the Scientific Study of Consciousness, and more directly in the creation of the Center for Consciousness Studies at the University of Arizona, of which Hameroff is now the director. The Center for Consciousness Studies hosts meetings on the study of consciousness every two years, as well as sponsoring seminars on consciousness theory.

Hameroff appeared as himself in the documentary film *What the #\$\*! Do  $\omega\Sigma(k)\pi\omega$ !?* (2004). He also participated in the first Beyond Belief conference, where his theories were sharply criticized by Lawrence Krauss, among others.<sup>[1]</sup>

Hameroff serves as producer, writer and scientific advisor to an independent feature film called Mindville. Mindville is a feature-length motion picture that combines live action with animation and effects to present a journey into the mysteries of human consciousness.

## Notes

- [1] "TSN: Session 4" (<http://thesciencenetwork.org/programs/beyond-belief-science-religion-reason-and-survival/session-4-1>). Thesciencenetwork.org. . Retrieved 2012-08-13.

## References

- Hameroff, S.(1987) *Ultimate Computing* Elsevier
- Penrose, R.(1989) *The Emperor's New Mind* Oxford University Press
- Grush, R. & Churchland, P. (1995) "Gap's in Penrose's Toilings," *Journal of Consciousness Studies*, **2**(1), pp. 10–29
- Penrose, R.(1994) *Shadows of the Mind* Oxford University Press
- Tegmark, M. (2000) "Importance of quantum coherence in brain processes," *Physical Reviews E*, **61**, pp. 4194–4206
- Hagan, S., Hameroff, S. & Tuszynski, J.(2002) "Quantum computation in brain microtubules? Decoherence and biological feasibility," *Physical Reviews E*, **65**, 061901.
- Hameroff, S. (2006) "Consciousness, neurobiology and quantum mechanics," In: *The Emerging Physics of Consciousness*, (Ed.) Tuszynski, J.


## Further reading

- Stuart Hameroff with Conrad Schneiker, *Ultimate Computing: Biomolecular Consciousness and Nanotechnology*, Elsevier-North Holland, 1987. This work predates the quantum Orch-OR hypothesis; still of interest. Online at author's site (<http://www.quantumconsciousness.org/ultimatecomputing.html>)
- Hameroff, Kaszniak, Scott, (eds), *Toward a Science of Consciousness*, MIT Press, ISBN 0-262-08249-7, LoC OP411.T68 1996. papers from the first Tucson conference on study of consciousness. Further volumes in the series exist.
- Roger Penrose, *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford, ISBN 0-19-853978-9, LoC Q335.P416 1994. This discusses the Orch-OR theory.

## External links

- Hameroff's "Quantum Consciousness" site (<http://www.quantumconsciousness.org/>)
- Center for consciousness studies homepage (<http://consciousness.arizona.edu/>)
- Quantum-Mind (<http://www.quantum-mind.co.uk>)

# Roger Penrose

<b>Roger Penrose</b>	
 <p style="text-align: center;">Roger Penrose, 2005</p>	
<b>Born</b>	8 August 1931 Colchester, Essex, England
<b>Residence</b>	<ul style="list-style-type: none"> <li>• United Kingdom</li> <li>• Canada (during WWII)</li> </ul>
<b>Nationality</b>	British
<b>Fields</b>	Mathematical physics
<b>Institutions</b>	<ul style="list-style-type: none"> <li>• Bedford College, London</li> <li>• St John's College, Cambridge</li> <li>• Princeton University</li> <li>• Syracuse University</li> <li>• King's College, London</li> <li>• Birkbeck, University of London</li> <li>• University of Oxford</li> <li>• Polish Academy of Sciences</li> </ul>
<b>Alma mater</b>	<ul style="list-style-type: none"> <li>• University of Cambridge</li> <li>• University College London</li> <li>• University College School</li> </ul>
<b>Doctoral advisor</b>	John A. Todd
<b>Other academic advisors</b>	William Hodge
<b>Doctoral students</b>	<ul style="list-style-type: none"> <li>• Tristan Needham</li> <li>• Richard Jozsa</li> <li>• Richard S. Ward</li> <li>• Andrew Hodges</li> <li>• Asghar Qadir</li> <li>• George Burnett-Stuart</li> <li>• Matthew Ginsberg</li> <li>• Adam Helfer</li> <li>• Lane P. Hughston</li> <li>• Peter Law</li> <li>• Claude LeBrun</li> <li>• Ross Moore</li> <li>• Duncan Stone</li> <li>• Erlangsen Bergstrom</li> <li>• Tim Poston</li> <li>• George Sparling</li> <li>• K. Paul Tod</li> </ul>



<b>Known for</b>	<ul style="list-style-type: none"> <li>• Twistor theory</li> <li>• Geometry of spacetime</li> <li>• Cosmic censorship</li> <li>• Weyl curvature hypothesis</li> <li>• Penrose inequalities</li> <li>• Penrose interpretation of Quantum Theory</li> <li>• Orch-OR</li> <li>• Moore–Penrose pseudoinverse</li> <li>• Newman–Penrose formalism</li> <li>• Penrose tiling</li> <li>• Penrose stairs</li> <li>• Penrose graphical notation</li> <li>• Schrödinger–Newton equations</li> </ul>
<b>Influences</b>	Dennis W. Sciama
<b>Influenced</b>	<ul style="list-style-type: none"> <li>• Michael Atiyah</li> <li>• Stuart Hameroff</li> </ul>
<b>Notable awards</b>	<ul style="list-style-type: none"> <li>• Wolf Prize (1988)</li> <li>• Dirac Medal (1989)</li> <li>• Copley Medal (2008)</li> </ul>
<b>Notes</b>	He is the brother of Jonathan Penrose and Oliver Penrose, and son of Lionel Penrose. He is the nephew of Roland Penrose.

**Sir Roger Penrose** OM FRS (born 8 August 1931), is an English mathematical physicist, recreational mathematician and philosopher. He is the Emeritus Rouse Ball Professor of Mathematics at the Mathematical Institute of the University of Oxford, as well as an Emeritus Fellow of Wadham College.

Penrose is internationally renowned for his scientific work in mathematical physics, in particular for his contributions to general relativity and cosmology. He has received a number of prizes and awards, including the 1988 Wolf Prize for physics, which he shared with Stephen Hawking for their contribution to our understanding of the universe.<sup>[1]</sup>

## Early life and academia

Born in Colchester, Essex, England, Roger Penrose is a son of Lionel S. Penrose and Margaret Leathes.<sup>[2]</sup> Penrose is the brother of mathematician Oliver Penrose and of chess Grandmaster Jonathan Penrose. Penrose attended University College School and University College, London, where he graduated with a first class degree in mathematics. In 1955, while still a student, Penrose reintroduced the E. H. Moore generalized matrix inverse, also known as the Moore–Penrose inverse,<sup>[3]</sup> after it had been reinvented by Arne Bjerhammar (1951). Penrose earned his Ph.D. at Cambridge (St John's College) in 1958, writing a thesis on "tensor methods in algebraic geometry" under algebraist and geometer John A. Todd. He devised and popularised the Penrose triangle in the 1950s, describing it as "impossibility in its purest form" and exchanged material with the artist M. C. Escher, whose earlier depictions of impossible objects partly inspired it. Escher's Waterfall, and Ascending and Descending were in turn inspired by Penrose. As reviewer Manjit Kumar puts it:

As a student in 1954, Penrose was attending a conference in Amsterdam when by chance he came across an exhibition of Escher's work. Soon he was trying to conjure up impossible figures of his own and discovered the tri-bar – a triangle that looks like a real, solid three-dimensional object, but isn't. Together with his father, a physicist and mathematician, Penrose went on to design a staircase that simultaneously loops up and down. An article followed and a copy was sent to Escher. Completing a cyclical flow of creativity, the Dutch master of geometrical illusions was inspired to produce his two masterpieces.<sup>[4]</sup>

In 1965, at Cambridge, Penrose proved that singularities (such as black holes) could be formed from the gravitational collapse of immense, dying stars.<sup>[5]</sup> This work was extended by Hawking to prove the Penrose–Hawking singularity

theorems.



Oil painting by Urs Schmid (1995) of a Penrose tiling using fat and thin rhombi.

In 1967, Penrose invented the twistor theory which maps geometric objects in Minkowski space into the 4-dimensional complex space with the metric signature (2,2). In 1969, he conjectured the cosmic censorship hypothesis. This proposes (rather informally) that the universe protects us from the inherent unpredictability of singularities (such as the one in the centre of a black hole) by hiding them from our view behind an event horizon. This form is now known as the "weak censorship hypothesis"; in 1979, Penrose formulated a stronger version called the "strong censorship hypothesis". Together with the BKL conjecture and issues of nonlinear stability, settling the censorship conjectures is one of the most important outstanding problems in general relativity. Also from 1979 dates

Penrose's influential Weyl curvature hypothesis on the initial conditions of the observable part of the Universe and the origin of the second law of thermodynamics.<sup>[6]</sup> Penrose and James Terrell independently realized that objects travelling near the speed of light will appear to undergo a peculiar skewing or rotation. This effect has come to be called the Terrell rotation or Penrose–Terrell rotation.<sup>[7][8]</sup>

Penrose is well known for his 1974 discovery of Penrose tilings, which are formed from two tiles that can only tile the plane nonperiodically, and are the first tilings to exhibit fivefold rotational symmetry. Penrose developed these ideas based on the article *Deux types fondamentaux de distribution statistique*<sup>[9]</sup> (1938; an English translation *Two Basic Types of Statistical Distribution*) by Czech geographer, demographer and statistician Jaromír Korčák. In 1984, such patterns were observed in the arrangement of atoms in quasicrystals.<sup>[10]</sup> Another noteworthy contribution is his 1971 invention of spin networks, which later came to form the geometry of spacetime in loop quantum gravity. He was influential in popularizing what are commonly known as Penrose diagrams (causal diagrams).

In 2004 Penrose released *The Road to Reality: A Complete Guide to the Laws of the Universe*, a 1,099-page book aimed at giving a comprehensive guide to the laws of physics. He has proposed a novel interpretation of quantum mechanics.<sup>[11]</sup> In 2010, Penrose reported possible evidence, based on concentric circles found in WMAP data of the CMB sky, of an earlier universe existing before the Big Bang of our own present universe.<sup>[12]</sup>

Penrose is the Francis and Helen Pentz Distinguished (visiting) Professor of Physics and Mathematics at Pennsylvania State University.<sup>[13]</sup> He is also a member of the *Astronomical Review* Editorial Board. Penrose is married to Vanessa Thomas, head of mathematics at Abingdon School,<sup>[14][15]</sup> with whom he has one son.<sup>[14]</sup> He has three sons from a previous marriage to American Joan Isabel Wedge, whom he married in 1959.

## Physics and consciousness

Penrose has written books on the connection between fundamental physics and human (or animal) consciousness. In *The Emperor's New Mind* (1989), he argues that known laws of physics are inadequate to explain the phenomenon of consciousness. Penrose proposes the characteristics this new physics may have and specifies the requirements for a bridge between classical and quantum mechanics (what he calls *correct quantum gravity*). Penrose uses a variant of Turing's halting theorem to demonstrate that a system can be deterministic without being algorithmic. (E.g., imagine a system with only two states, ON and OFF. If the system's state is ON if a given Turing machine halts, and OFF if the Turing machine does not halt, then the system's state is completely determined by the Turing machine, however there is no algorithmic way to determine whether the Turing machine stops.)

Penrose believes that such deterministic yet non-algorithmic processes may come in play in the quantum mechanical wave function reduction, and may be harnessed by the brain. He argues that the present computer is unable to have intelligence because it is an algorithmically deterministic system. He argues against the viewpoint that the rational processes of the mind are completely algorithmic and can thus be duplicated by a sufficiently complex computer. This contrasts with supporters of strong artificial intelligence, who contend that thought can be simulated algorithmically. He bases this on claims that consciousness transcends formal logic because things such as the insolubility of the halting problem and Gödel's incompleteness theorem prevent an algorithmically based system of logic from reproducing such traits of human intelligence as mathematical insight. These claims were originally espoused by the philosopher John Lucas of Merton College, Oxford.

The Penrose/Lucas argument about the implications of Gödel's incompleteness theorem for computational theories of human intelligence has been widely criticized by mathematicians, computer scientists and philosophers, and the consensus among experts in these fields seems to be that the argument fails, though different authors may choose different aspects of the argument to attack.<sup>[16]</sup> Marvin Minsky, a leading proponent of artificial intelligence, was particularly critical, stating that Penrose "tries to show, in chapter after chapter, that human thought cannot be based on any known scientific principle." Minsky's position is exactly the opposite - he believes that humans are, in fact, machines, whose functioning, although complex, is fully explainable by current physics. Minsky maintains that "one can carry that quest [for scientific explanation] too far by only seeking new basic principles instead of attacking the real detail. This is what I see in Penrose's quest for a new basic principle of physics that will account for consciousness."<sup>[17]</sup>

Penrose responded to criticism of *The Emperor's New Mind* with his follow up 1994 book *Shadows of the Mind*, and in 1997 with *The Large, the Small and the Human Mind*. In those works, he also combined his observations with that of anesthesiologist Stuart Hameroff.

Penrose and Hameroff have argued that consciousness is the result of quantum gravity effects in microtubules, which they dubbed Orch-OR (orchestrated objective reduction). Max Tegmark, in a paper in *Physical Review E*,<sup>[18]</sup> calculated that the time scale of neuron firing and excitations in microtubules is slower than the decoherence time by a factor of at least 10,000,000,000. The reception of the paper is summed up by this statement in Tegmark's support: "Physicists outside the fray, such as IBM's John A. Smolin, say the calculations confirm what they had suspected all along. 'We're not working with a brain that's near absolute zero. It's reasonably unlikely that the brain evolved



Prof. Penrose at a conference.

quantum behavior".<sup>[19]</sup> Tegmark's paper has been widely cited by critics of the Penrose–Hameroff position.

In their reply to Tegmark's paper, also published in *Physical Review E*, the physicists Scott Hagan, Jack Tuszynski and Hameroff<sup>[20][21]</sup> claimed that Tegmark did not address the Orch-OR model, but instead a model of his own construction. This involved superpositions of quanta separated by 24 nm rather than the much smaller separations stipulated for Orch-OR. As a result, Hameroff's group claimed a decoherence time seven orders of magnitude greater than Tegmark's, but still well short of the 25 ms required if the quantum processing in the theory was to be linked to the 40 Hz gamma synchrony, as Orch-OR suggested. To bridge this gap, the group made a series of proposals. It was supposed that the interiors of neurons could alternate between liquid and gel states. In the gel state, it was further hypothesized that the water electrical dipoles are oriented in the same direction, along the outer edge of the microtubule tubulin subunits. Hameroff et al. proposed that this ordered water could screen any quantum coherence within the tubulin of the microtubules from the environment of the rest of the brain. Each tubulin also has a tail extending out from the microtubules, which is negatively charged, and therefore attracts positively charged ions. It is suggested that this could provide further screening. Further to this, there was a suggestion that the microtubules could be pumped into a coherent state by biochemical energy.

Finally, it is suggested that the configuration of the microtubule lattice might be suitable for quantum error correction, a means of holding together quantum coherence in the face of environmental interaction. In the last decade, some researchers who are sympathetic to Penrose's ideas have proposed an alternative scheme for quantum processing in microtubules based on the interaction of tubulin tails with microtubule-associated proteins, motor proteins and presynaptic scaffold proteins. These proposed alternative processes have the advantage of taking place within Tegmark's time to decoherence.

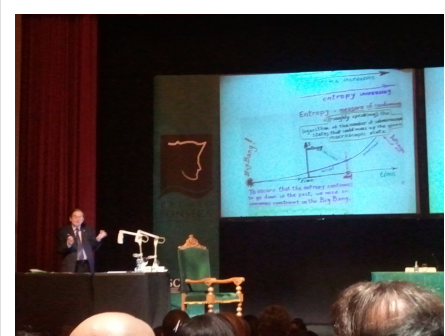
In 2011 W. Christensen<sup>[22]</sup> argued that the universe can be shown to be conscious via a cosmological model based on Maxwell's demon and information theory.

Hameroff, in a lecture in part of a Google Tech talks series exploring quantum biology, gave an overview of current research in the area, and responded to subsequent criticisms of the Orch-OR model.<sup>[23]</sup> In addition to this, a recent 2011 paper by Roger Penrose and Stuart Hameroff gives an updated model of their Orch-OR theory, in light of criticisms, and discusses the place of consciousness within the universe.<sup>[24]</sup>

Phillip Tetlow, although himself supportive of Penrose's views, acknowledges that Penrose's ideas about the human thought process are at present a minority view in scientific circles, citing Minsky's criticisms and quoting science journalist Charles Seife's description of Penrose as "one of a handful of scientists" who believe that the nature of consciousness suggests a quantum process.<sup>[19]</sup>

## Religious views

Penrose does not hold to any religious doctrine,<sup>[25]</sup> and refers to himself as an atheist.<sup>[26]</sup> In the film *A Brief History of Time*, he said, "I think I would say that the universe has a purpose, it's not somehow just there by chance ... some people, I think, take the view that the universe is just there and it runs along – it's a bit like it just sort of computes, and we happen somehow by accident to find ourselves in this thing. But I don't think that's a very fruitful or helpful way of looking at the universe, I think that there is something much deeper about it."<sup>[27]</sup> Penrose is a Distinguished Supporter of the British Humanist Association.



Roger Penrose in the University of Santiago de Compostela to receive the Fonseca Prize.

## Awards and honours

Penrose has been awarded many prizes for his contributions to science. He was elected a Fellow of the Royal Society of London in 1972. In 1975, Stephen Hawking and Penrose were jointly awarded the Eddington Medal of the Royal Astronomical Society. In 1985, he was awarded the Royal Society Royal Medal. Along with Stephen Hawking, he was awarded the prestigious Wolf Foundation Prize for Physics in 1988. In 1989 he was awarded the Dirac Medal and Prize of the British Institute of Physics. In 1990 Penrose was awarded the Albert Einstein Medal for outstanding work related to the work of Albert Einstein by the Albert Einstein Society. In 1991, he was awarded the Naylor Prize of the London Mathematical Society. From 1992 to 1995 he served as President of the International Society on General Relativity and Gravitation<sup>[28]</sup>. In 1994, Penrose was knighted for services to science.<sup>[29]</sup> In the same year he was also awarded an Honorary Degree (Doctor of Science) by the University of Bath.<sup>[30]</sup> In 1998, he was elected Foreign Associate of the United States National Academy of Sciences. In 2000 he was appointed to the Order of Merit. In 2004 he was awarded the De Morgan Medal for his wide and original contributions to mathematical physics. To quote the citation from the London Mathematical Society:

His deep work on General Relativity has been a major factor in our understanding of black holes. His development of Twistor Theory has produced a beautiful and productive approach to the classical equations of mathematical physics. His tilings of the plane underlie the newly discovered quasi-crystals.<sup>[31]</sup>

In 2005 Penrose was awarded an honorary doctorate by Warsaw University and Katholieke Universiteit Leuven (Belgium), and in 2006 by the University of York. In 2008 Penrose was awarded the Copley Medal. He is also a Distinguished Supporter of the British Humanist Association and one of the patrons of the Oxford University Scientific Society. In 2011, Penrose was awarded the Fonseca Prize by the University of Santiago de Compostela. In 2012 Penrose was awarded the Richard R. Ernst Medal by ETH Zürich for his contributions to science and strengthening the connection between science and society.

## Works

- *Techniques of Differential Topology in Relativity* (1972, ISBN 0-89871-005-7)
- *Spinors and Space-Time: Volume 1, Two-Spinor Calculus and Relativistic Fields* (with Wolfgang Rindler, 1987) ISBN 0-521-33707-0 (paperback)
- *Spinors and Space-Time: Volume 2, Spinor and Twistor Methods in Space-Time Geometry* (with Wolfgang Rindler, 1988) (reprint), ISBN 0-521-34786-6 (paperback)
- *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics* (1989, ISBN 0-14-014534-6 (paperback); it received the Rhone-Poulenc science book prize in 1990)
- *Shadows of the Mind: A Search for the Missing Science of Consciousness* (1994, ISBN 0-19-853978-9 (hardback))
- *The Nature of Space and Time* (with Stephen Hawking, 1996, ISBN 0-691-03791-4 (hardback), ISBN 0-691-05084-8 (paperback))
- *The Large, the Small, and the Human Mind* (with Abner Shimony, Nancy Cartwright, and Stephen Hawking, 1997, ISBN 0-521-56330-5 (hardback), ISBN 0-521-65538-2 (paperback), Canto edition: ISBN 0-521-78572-3)
- *White Mars or, The Mind Set Free* (with Brian W. Aldiss, 1999, ISBN 978-0-316-85243-2 (hardback))
- *The Road to Reality: A Complete Guide to the Laws of the Universe* (2004, ISBN 0-224-04447-8 (hardcover), ISBN 0-09-944068-7 (paperback))



Roger Penrose during a lecture



- *Cycles of Time: An Extraordinary New View of the Universe* (Bodley Head (23 Sep 2010) ISBN 978-0-224-08036-1)

Penrose also wrote forewords to *Quantum Aspects of Life* and Zee's book *Fearful Symmetry*.

## References

- [1] Penrose, R (2005). *The Road to Reality: A Complete guide to the Laws of the Universe*. Vintage Books. ISBN 0-09-944068-7
  - [2] Penrose and his father shared mathematical concepts with Dutch graphic artist M. C. Escher which were incorporated into a lot of pieces, including Waterfall, which is based on the 'Penrose triangle', and Up and Down.
  - [3] Penrose, R. "A Generalized Inverse for Matrices" *Proc. Cambridge Phil. Soc.* 51, 406–413, 1955
  - [4] Cycles of Time: An Extraordinary New View of the Universe by Roger Penrose – review | Books | The Guardian (<http://www.guardian.co.uk/books/2010/oct/16/cycles-time-roger-penrose-review>)
  - [5] Ferguson, 1991: 66
  - [6] R. Penrose (1979). "Singularities and Time-Asymmetry". In S. W. Hawking and W. Israel. *General Relativity: An Einstein Centenary Survey*. Cambridge University Press. pp. 581–638.
  - [7] Terrell, James (1959). "Invisibility of the Lorentz Contraction". *Physical Review* **116** (4): 1041–1045. Bibcode 1959PhRv..116.1041T. doi:10.1103/PhysRev.116.1041
  - [8] Penrose, Roger (1959). "The Apparent Shape of a Relativistically Moving Sphere". *Proceedings of the Cambridge Philosophical Society* **55**: 137–139. Bibcode 1959PCPS...55..137P. doi:10.1017/S0305004100033776
  - [9] Jaromír Korčák (1938): Deux types fondamentaux de distribution statistique. Prague, Comité d'organisation, Bull. de l'Institute Int'l de Statistique, vol. 3, pp. 295–299.
  - [10] Steinhardt, Paul (1996). "New perspectives on forbidden symmetries, quasicrystals, and Penrose tilings" (<http://www.pnas.org/content/93/25/14267.full>). *PNAS* **93** (25): 14267–14270. Bibcode 1996PNAS...9314267S. doi:10.1073/pnas.93.25.14267. PMC 34472. PMID 8962037.
  - [11] "If an Electron Can Be in Two Places at once, Why Can't You?" ([http://discovermagazine.com/2005/jun/cover/article\\_view?b\\_start:int=0&-C=](http://discovermagazine.com/2005/jun/cover/article_view?b_start:int=0&-C=)). . Retrieved 2008-10-27.
  - [12] Gurzadyan, V.G.; Penrose, R. (2010). "Concentric circles in WMAP data may provide evidence of violent pre-Big-Bang activity". arXiv:1011.3706 [astro-ph.CO].
  - [13] "Dr. Roger Penrose at Penn State University" ([http://www.phys.psu.edu/people/display/index.html?person\\_id=233&mode=contact](http://www.phys.psu.edu/people/display/index.html?person_id=233&mode=contact)). . Retrieved 2007-07-09.
  - [14] "The Peter & Patricia Gruber Foundation, St. Thomas US Virgin Islands - Grants and International Awards" ([http://www.gruberprizes.org/SelectingRecipients/SelectionAdvisoryBoard\\_Bio.php?id=4](http://www.gruberprizes.org/SelectingRecipients/SelectionAdvisoryBoard_Bio.php?id=4)). Gruberprizes.org. 1931-08-08. . Retrieved 2012-08-13.
  - [15] "Vanessa Penrose | Abingdon School" ([http://www.abingdon.org.uk/vanessa\\_penrose/](http://www.abingdon.org.uk/vanessa_penrose/)). Abingdon.org.uk. 2012-07-06. . Retrieved 2012-08-13.
  - [16] Criticism of the Lucas/Penrose argument that intelligence can not be entirely algorithmic:
    - MindPapers: 6.1b. Godelian arguments (<http://consc.net/mindpapers/6.1b>)
    - References for Criticisms of the Gödelian Argument (<http://users.ox.ac.uk/~jrlucas/Godel/referenc.html>)\*
    - Boolos, George, et al. 1990. *An Open Peer Commentary on The Emperor's New Mind*. Behavioral and Brain Sciences 13 (4) 655.
    - Davis, Martin 1993. *How subtle is Gödel's theorem? More on Roger Penrose*. Behavioral and Brain Sciences, 16, 611-612. Online version at Davis' faculty page at <http://cs.nyu.edu/cs/faculty/davism/>
    - Feferman, Solomon (1996). "Penrose's Gödelian argument". *PSYCHE* **2**: 21–32. CiteSeerX: 10.1.1.130.7027 (<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.130.7027>).
    - Krajewski, Stanislaw 2007. *On Gödel's Theorem and Mechanism: Inconsistency or Unsoundness is Unavoidable in any Attempt to 'Out-Gödel' the Mechanist*. Fundamenta Informaticae 81, 173–181. Reprinted in Topics in Logic, Philosophy and Foundations of Mathematics and Computer Science: In Recognition of Professor Andrzej Grzegorzczak (2008), p. 173 (<http://books.google.com/books?id=0jSS-3B106cC&lpg=PP1&pg=PA173#v=onepage&q&f=false>)
    - LaForte, Geoffrey, Patrick J. Hayes, and Kenneth M. Ford 1998. *Why Gödel's Theorem Cannot Refute Computationalism* (<http://www.cs.uwf.edu/~glaforte/papers/whyGodel.ps>). Artificial Intelligence, 104:265–286.
    - Lewis, David K. 1969. *Lucas against mechanism* ([http://www2.units.it/etica/2003\\_1/7\\_monographica.doc](http://www2.units.it/etica/2003_1/7_monographica.doc)). Philosophy 44 231–233.
    - Putnam, Hilary 1995. *Review of Shadows of the Mind*. In Bulletin of the American Mathematical Society 32, 370–373 (also see Putnam's less technical criticisms in his New York Times review (<http://www.nytimes.com/books/97/04/27/nnp/17540.html>))
- Sources that indicate Penrose's argument is generally rejected:
- Bringsford, S. and Xiao, H. 2000. *A Refutation of Penrose's Gödelian Case Against Artificial Intelligence* (<http://kryten.mm.rpi.edu/refute.penrose.pdf>). Journal of Experimental and Theoretical Artificial Intelligence 12: 307–329. The authors write that it is "generally agreed" that Penrose "failed to destroy the computational conception of mind."
  - In an article at <http://www.mth.kcl.ac.uk/~llandau/Homepage/Math/penrose.html> L.J. Landau at the Mathematics Department of King's College London writes that "Penrose's argument, its basis and implications, is rejected by experts in the fields which it touches."
- Sources that also note that different sources attack different points of the argument:

- Princeton Philosophy professor John Burgess writes in *On the Outside Looking In: A Caution about Conservativeness* (<http://www.princeton.edu/~jburgess/Montreal.doc>) (published in Kurt Gödel: Essays for his Centennial, with the following comments found on pp. 131–132 (<http://books.google.com/books?id=83Att6BsJ4C&lpq=PP1&pg=PA131#v=onepage&q&f=false>)) that "the consensus view of logicians today seems to be that the Lucas–Penrose argument is fallacious, though as I have said elsewhere, there is at least this much to be said for Lucas and Penrose, that logicians are not unanimously agreed as to where precisely the fallacy in their argument lies. There are at least three points at which the argument may be attacked."
  - Dershowitz, Nachum 2005. *The Four Sons of Penrose* (<http://www.cs.tau.ac.il/~nachumd/papers/FourSonsOfPenrose.pdf>), in *Proceedings of the Eleventh Conference on Logic Programming for Artificial Intelligence and Reasoning (LPAR; Jamaica)*, G. Sutcliffe and A. Voronkov, eds., Lecture Notes in Computer Science, vol. 3835, Springer-Verlag, Berlin, pp. 125–138.
- [17] Marvin Minsky. "Conscious Machines." Machinery of Consciousness, Proceedings, National Research Council of Canada, 75th Anniversary Symposium on Science in Society, June 1991.
- [18] Tegmark, Max. 2000. "The importance of quantum decoherence in brain processes". Physical Review E. vol 61. pp. 4194–4206.
- [19] Tetlow, Philip (2007). *The Web's Awake: An Introduction to the Field of Web Science and the Concept of Web Life* ([http://books.google.com/?id=3mPI9rUuhJ8C&printsec=frontcover&dq=penrose+\"thought+process\"&q=penrose](http://books.google.com/?id=3mPI9rUuhJ8C&printsec=frontcover&dq=penrose+\) ). Hoboken, NJ: John Wiley & Sons. p. 166. ISBN 978-0-470-13794-9. .
- [20] Hagan, S., Hameroff, S., and Tuszynski, J. (2002). "Quantum Computation in Brain Microtubules? Decoherence and Biological Feasibility". *Physical Review E* **65**: 061901. arXiv:quant-ph/0005025. Bibcode 2002PhRvE..65f1901H. doi:10.1103/PhysRevE.65.061901.
- [21] Hameroff, S. (2006). "Consciousness, Neurobiology and Quantum Mechanics". In Tuszynski, Jack. *The Emerging Physics of Consciousness*. Springer. pp. 193–253
- [22] Sir Roger Penrose & Dr. Stuart Hameroff (2011). "*Consciousness and the Universe: Quantum Physics, Evolution, Brain and Mind*," Chapter XIV *Does the Universe have Cosmological Memory? If so does this imply Cosmic Consciousness?*. Oxford University Press.
- [23] "Clarifying the Tubulin bit/qubit - Defending the Penrose-Hameroff Orch OR Model (Quantum Biology)" (<http://www.youtube.com/watch?v=LXFFbxoHp3s>). YouTube. 2010-10-22. . Retrieved 2012-08-13.
- [24] Journal of Cosmology.com (1992-07-04). "Journal of Cosmology" (<http://www.quantumconsciousness.org/Cosmology160.html>). Quantumconsciousness.org. . Retrieved 2012-08-13.
- [25] Harris, Sam. "Letter to A Christian Nation" ([http://www.samharris.org/site/book\\_letter\\_to\\_christian\\_nation/](http://www.samharris.org/site/book_letter_to_christian_nation/)). *SamHarrisOrg*. . Retrieved 5 June 2010. Quoting Penrose's blurb for Harris's book *Letter to a Christian Nation*.
- [26] "Big Bang follows Big Bang follows Big Bang" (<http://news.bbc.co.uk/1/hi/9032000/9032626.stm>). *BBC News*. 25 September 2010. . Retrieved 1 Dec 2010.
- [27] See A Brief History of Time (<http://video.google.com/videoplay?docid=-5382609196671955257#>), quote starts at about 1:12:43 in the video.
- [28] <http://grg.maths.qmul.ac.uk/grgsoc/>
- [29] Official announcement knighthood. (<http://www.gazettes-online.co.uk/ViewPDF.aspx?pdf=53696&geotype=London&gpn=2&type=ArchivedSupplementPage&exact=Roger Penrose>) The London Gazette 11 June 1994.
- [30] "Honorary Graduates 1989 to present" (<http://www.bath.ac.uk/ceremonies/hongrads/>). *bath.ac.uk*. University of Bath. . Retrieved 18 February 2012.
- [31] London Mathematical Society ([http://www.lms.ac.uk/activities/prizes\\_com/citations04.html](http://www.lms.ac.uk/activities/prizes_com/citations04.html))

## Further reading

- Ferguson, Kitty (1991). *Stephen Hawking: Quest For A Theory of Everything*. Franklin Watts. ISBN 0-553-29895-X.
- Misner, Charles; Thorne, Kip S. & Wheeler, John Archibald (1973). *Gravitation*. San Francisco: W. H. Freeman. ISBN 0-7167-0344-0. (See *Box 34.2*.)

## External links

- Dangerous Knowledge (<http://www.youtube.com/watch?v=Cw-zNRNcF90>) – Penrose was one of the principal interviewees in a BBC documentary about the mathematics of infinity directed by David Malone
- Penrose's new theory "Aeons Before the Big Bang?":
  - Original 2005 lecture: "Before the Big Bang? A new perspective on the Weyl curvature hypothesis" (<http://www.newton.ac.uk/webseminars/pg+ws/2005/gmr/gmrw04/1107/penrose/frames.html>) (Isaac Newton Institute for Mathematical Sciences, Cambridge, Nov 11, 2005).
  - Original publication: "Before the Big Bang: an outrageous new perspective and its implications for particle physics". (<http://accelconf.web.cern.ch/AccelConf/e06/PAPERS/THESPA01.PDF>) *Proceedings of*

- EPAC 2006*. Edinburgh. 2759–2762 (cf. also Hill, C.D. & Nurowski, P. (2007) "On Penrose's 'Before the Big Bang' ideas" (<http://arxiv.org/pdf/0710.3879v2>). Ithaca)
- Revised 2009 lecture: "Aeons Before the Big Bang?" (<http://smartech.gatech.edu/handle/1853/27632>) (Georgia Institute of Technology, Center for Relativistic Astrophysics)
  - BBC interview (<http://www.youtube.com/watch?v=pElj9zcLzp0>) on the new theory
  - Roger Penrose on The Forum ([http://www.bbc.co.uk/worldservice/programmes/2009/03/000000\\_forum.shtml](http://www.bbc.co.uk/worldservice/programmes/2009/03/000000_forum.shtml))
  - Penrose on sidestepping reason (<http://www.youtube.com/watch?v=xiYDc1LA0I4&feature=user>)
  - O'Connor, John J.; Robertson, Edmund F., "Roger Penrose" (<http://www-history.mcs.st-andrews.ac.uk/Biographies/Penrose.html>), *MacTutor History of Mathematics archive*, University of St Andrews.
  - Shaking Up Foundations of Math: Roger Penrose on Kurt Gödel's Groundbreaking Work ([http://simplycharly.com/godel/roger\\_penrose\\_godel\\_interview.html](http://simplycharly.com/godel/roger_penrose_godel_interview.html))
  - Hilary Putnam's review of Penrose's 'Shadows of the Mind' claiming that Penrose's use of Gödel's Incompleteness Theorem is fallacious (<http://www.ams.org/bull/pre-1996-data/199507/199507015.pdf>)
    - **Beyond the Doubting of a Shadow: A Reply to Commentaries on Shadows of the Mind** (<http://web.archive.org/web/20080618195657/http://psyche.csse.monash.edu.au/v2/psyche-2-23-penrose.html>)
  - Penrose Tiling found in Islamic Architecture (<http://www.sciencenews.org/articles/20070224/mathtrek.asp>)
  - Two theories for the formation of quasicrystals resembling Penrose tilings ([http://www.sciencenews.org/sn\\_arch/10\\_12\\_96/bob1.htm](http://www.sciencenews.org/sn_arch/10_12_96/bob1.htm))
  - Tegmark, Max. 2000. "The importance of quantum decoherence in brain processes (<http://www.arxiv.org/abs/quant-ph/9907009>)". *Physical Review E*. vol 61. pp. 4194–4206.
    - "Biological feasibility of quantum states in the brain (<http://www.quantumconsciousness.org/pdfs/decoherence.pdf>)" – (a disputation of Tegmark's result by Hagan, Hameroff, and Tuszynski)
      - Tegmark's rejoinder to Hagan *et al.* (<http://space.mit.edu/home/tegmark/brain.html>)
  - "Toilet Paper Plagiarism" ([http://web.archive.org/web/20050312084035/http://www.parascope.com/articles/slips/fs\\_151.htm](http://web.archive.org/web/20050312084035/http://www.parascope.com/articles/slips/fs_151.htm)) – D. Trull about Penrose's lawsuit concerning the use of his Penrose tilings on toilet paper
  - Roger Penrose: A Knight on the tiles (Plus magazine) (<http://plus.maths.org/issue18/features/penrose/index.html>)
  - Penrose's Gifford Lecture biography (<http://www.giffordlectures.org/Author.asp?AuthorID=254>)
  - Quantum-Mind (<http://www.quantum-mind.co.uk>)
  - Audio: Roger Penrose in conversation on the BBC World Service discussion show ([http://www.bbc.co.uk/worldservice/documentaries/2009/04/090427\\_theforum\\_260409.shtml](http://www.bbc.co.uk/worldservice/documentaries/2009/04/090427_theforum_260409.shtml))

#### *The Forum*

- Roger Penrose speaking about Hawking's new book on Premier Christian Radio (<http://www.premierradio.org.uk/listen/ondemand.aspx?mediaid={320D8898-A8F0-4433-8934-D64DDEB8A21C}>)



# Holography

**Holography** is a technique which enables three-dimensional images to be made. It involves the use of a laser, interference, diffraction, light intensity recording and suitable illumination of the recording. The image changes as the position and orientation of the viewing system changes in exactly the same way as if the object were still present, thus making the image appear three-dimensional.

The holographic recording itself is not an image; it consists of an apparently random structure of either varying intensity, density or profile.

## Overview and history

The Hungarian-British physicist Dennis Gabor (Hungarian name: Gábor Dénes),<sup>[1][2]</sup> was awarded the Nobel Prize in Physics in 1971 "for his invention and development of the holographic method".<sup>[3]</sup> His work, done in the late 1940s, built on pioneering work in the field of X-ray microscopy by other scientists including Mieczysław Wolfke in 1920 and WL Bragg in 1939.<sup>[4]</sup> The discovery was an unexpected result of research into improving electron microscopes at the British

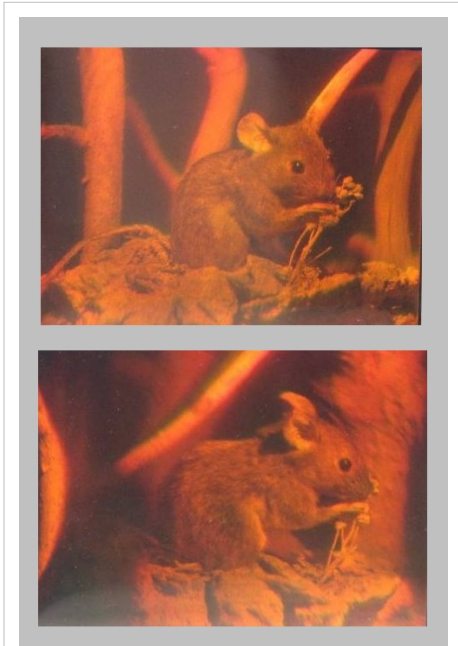
Thomson-Houston Company in Rugby, England, and the company filed a patent in December 1947 (patent GB685286). The technique as originally invented is still used in electron microscopy, where it is known as electron holography, but optical holography did not really advance until the development of the laser in 1960. The word *holography* comes from the Greek words ὅλος (*hólos*; "whole") and γραφή (*grafē*; "writing" or "drawing").

The development of the laser enabled the first practical optical holograms that recorded 3D objects to be made in 1962 by Yuri Denisyuk in the Soviet Union<sup>[5]</sup> and by Emmett Leith and Juris Upatnieks at the University of Michigan, USA.<sup>[6]</sup> Early holograms used silver halide photographic emulsions as the recording medium. They were not very efficient as the grating produced absorbed much of the incident light. Various methods of converting the variation in transmission to a variation in refractive index (known as "bleaching") were developed which enabled much more efficient holograms to be produced.<sup>[7][8][9]</sup>

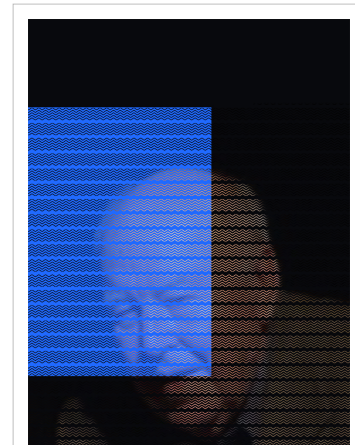
Several types of holograms can be made. Transmission holograms, such as those produced by Leith and Upatnieks, are viewed by shining laser light through them and looking at the reconstructed image from the side of the hologram opposite the source.<sup>[10]</sup> A later refinement, the "rainbow transmission" hologram, allows more convenient illumination by white light rather than by lasers.<sup>[11]</sup> Rainbow holograms are commonly used for security and authentication, for example, on credit cards and product packaging.<sup>[12]</sup>

Another kind of common hologram, the reflection or Denisyuk hologram, can also be viewed using a white-light illumination source on the same side of the hologram as the viewer and is the type of hologram normally seen in holographic displays. They are also capable of multicolour-image reproduction.<sup>[13]</sup>

Specular holography is a related technique for making three-dimensional images by controlling the motion of specularities on a two-dimensional surface.<sup>[14]</sup> It works by reflectively or refractively manipulating bundles of light



Two photographs of a single hologram taken from different viewpoints



Portrait of Yuri Denisyuk, by Dieter Jung

rays, whereas Gabor-style holography works by diffractively reconstructing wavefronts.

Most holograms produced are of static objects but systems for displaying changing scenes on a holographic volumetric display are now being developed.<sup>[15][16][17]</sup>

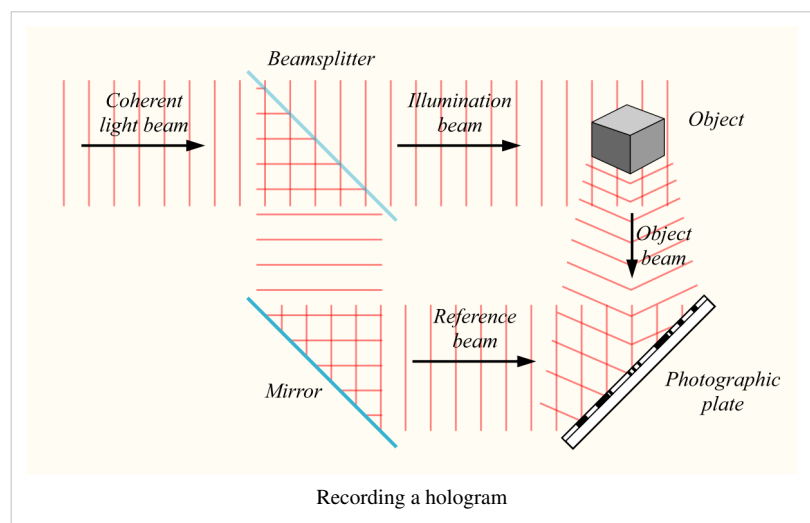
Holograms can also be used to store, retrieve, and process information optically.<sup>[18]</sup>

In its early days, holography required high-power expensive lasers, but nowadays, mass-produced low-cost semi-conductor or diode lasers, such as those found in millions of DVD recorders and used in other common applications, can be used to make holograms and have made holography much more accessible to low-budget researchers, artists and dedicated hobbyists.

It was thought that it would be possible to use X-rays to make holograms of molecules and view them using visible light. However, X-ray holograms have not been created to date.<sup>[19]</sup>

## How holography works

Holography is a technique that enables a light field, which is generally the product of a light source scattered off objects, to be recorded and later reconstructed when the original light field is no longer present, due to the absence of the original objects.<sup>[20]</sup> Holography can be thought of as somewhat similar to sound recording, whereby a sound field created by vibrating matter like musical instruments or vocal cords, is encoded in such a way that it can be reproduced later, without the presence of the original vibrating matter.



## Laser

Holograms are recorded using a flash of light that illuminates a scene and then imprints on a recording medium, much in the way a photograph is recorded. In addition, however, part of the light beam must be shone directly onto the recording medium - this second light beam is known as the reference beam. A hologram requires a laser as the sole light source. Lasers can be precisely controlled and have a fixed wavelength, unlike sunlight or light from conventional sources, which contain many different wavelengths.

To prevent external light from interfering, holograms are usually taken in darkness, or in low level light of a different colour from the laser light used in making the hologram.

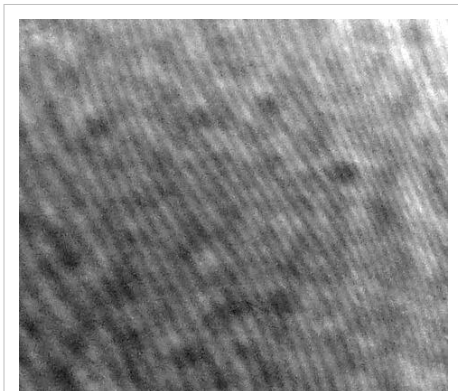
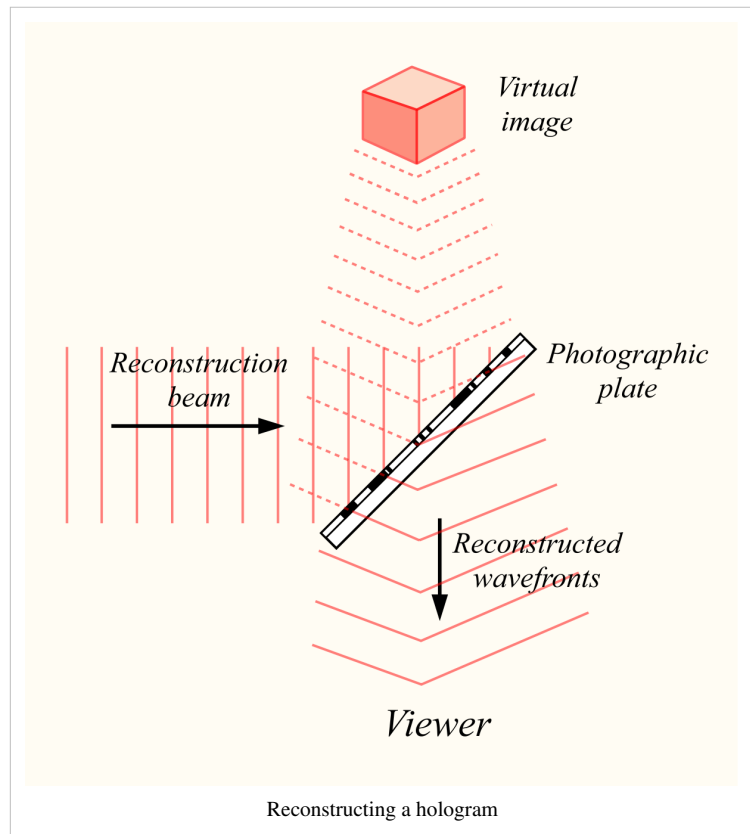
Holography requires a specific exposure time (just like photography), which can be controlled using a shutter, or by electronically timing the laser

### Apparatus

A hologram can be made by shining part of the light beam directly onto the recording medium, and the other part onto the object in such a way that some of the scattered light falls onto the recording medium.

A more flexible arrangement for recording a hologram requires the laser beam to be aimed through a series of elements that change it in different ways. The first element is a beam splitter that divides the beam into two identical beams, each aimed in different directions:

- One beam (known as the *illumination* or *object beam*) is spread using lenses and directed onto the scene using mirrors. Some of the light scattered (reflected) from the scene then falls onto the recording medium.
- The second beam (known as the *reference beam*) is also spread through the use of lenses, but is directed so that it doesn't come in contact with the scene, and instead travels directly onto the recording medium.



Close-up photograph of a hologram's surface. The object in the hologram is a toy van. It is no more possible to discern the subject of a hologram from this pattern than it is to identify what music has been recorded by looking at a CD surface. Note that the hologram is described by the speckle pattern, rather than the "wavy" line pattern.

Several different materials can be used as the recording medium. One of the most common is a film very similar to photographic film (silver halide photographic emulsion), but with a much higher concentration of light-reactive grains, making it capable of the much higher resolution that holograms require. A layer of this recording medium (e.g. silver halide) is attached to a transparent substrate, which is commonly glass, but may also be plastic.

## Process

When the two laser beams reach the recording medium, their light waves intersect and interfere with each other. It is this interference pattern that is imprinted on the recording medium. The pattern itself is seemingly random, as it represents the way in which the scene's light *interfered* with the original light source — but not the original light source itself. The interference pattern can be considered an encoded version of the scene, requiring a particular key — the original light source — in order to view its contents.

This missing key is provided later by shining a laser, identical to the one used to record the hologram, onto the developed film. When this beam illuminates the hologram, it is diffracted by the hologram's surface pattern. This produces a light field identical to the one originally produced by the scene and scattered onto the hologram. The image this effect produces in a person's retina is known as a virtual image.

## Holography vs. photography

Holography may be better understood via an examination of its differences from ordinary photography:

- A hologram represents a recording of information regarding the light that came from the original scene as scattered in a range of directions rather than from only one direction, as in a photograph. This allows the scene to be viewed from a range of different angles, as if it were still present.
- A photograph can be recorded using normal light sources (sunlight or electric lighting) whereas a laser is required to record a hologram.
- A lens is required in photography to record the image, whereas in holography, the light from the object is scattered directly onto the recording medium.
- A holographic recording requires a second light beam (the reference beam) to be directed onto the recording medium.
- A photograph can be viewed in a wide range of lighting conditions, whereas holograms can only be viewed with very specific forms of illumination.
- When a photograph is cut in half, each piece shows half of the scene. When a hologram is cut in half, the whole scene can still be seen in each piece. This is because, whereas each point in a photograph only represents light scattered from a single point in the scene, *each point* on a holographic recording includes information about light scattered from *every point* in the scene. Think of viewing a street outside your house through a 4 ft x 4 ft window, and then through a 2 ft x 2 ft window. You can see all of the same things through the smaller window (by moving your head to change your viewing angle), but you can see more *at once* through the 4 ft window.
- A photograph is a two-dimensional representation that can only reproduce a rudimentary three-dimensional effect, whereas the reproduced viewing range of a hologram adds many more depth perception cues that were present in the original scene. These cues are recognized by the human brain and translated into the same perception of a three-dimensional image as when the original scene might have been viewed.
- A photograph clearly maps out the light field of the original scene. The developed hologram's surface consists of a very fine, seemingly random pattern, which appears to bear no relationship to the scene it recorded.

## Physics of holography

For a better understanding of the process, it is necessary to understand interference and diffraction. Interference occurs when one or more wavefronts are superimposed. Diffraction occurs whenever a wavefront encounters an object. The process of producing a holographic reconstruction is explained below purely in terms of interference and diffraction. It is somewhat simplified but is accurate enough to provide an understanding of how the holographic process works.

For those unfamiliar with these concepts, it is worthwhile to read the respective articles before reading further in this article.

### Plane wavefronts

A diffraction grating is a structure with a repeating pattern. A simple example is a metal plate with slits cut at regular intervals. A light wave incident on a grating is split into several waves; the direction of these diffracted waves is determined by the grating spacing and the wavelength of the light.

A simple hologram can be made by superimposing two plane waves from the same light source on a holographic recording medium. The two waves interfere giving a straight line fringe pattern whose intensity varies sinusoidally across the medium. The spacing of the fringe pattern is determined by the angle between the two waves, and on the wavelength of the light.

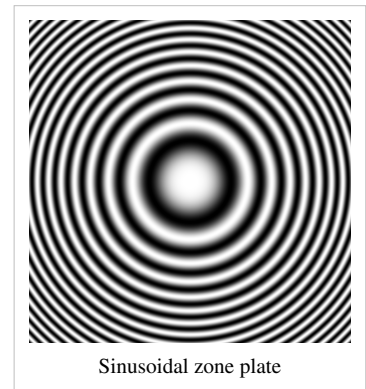
The recorded light pattern is a diffraction grating. When it is illuminated by only one of the waves used to create it, it can be shown that one of the diffracted waves emerges at the same angle as that at which the second wave was originally incident so that the second wave has been 'reconstructed'. Thus, the recorded light pattern is a holographic recording as defined above.

### Point sources

If the recording medium is illuminated with a point source and a normally incident plane wave, the resulting pattern is a sinusoidal zone plate which acts as a negative Fresnel lens whose focal length is equal to the separation of the point source and the recording plane.

When a plane wavefront illuminates a negative lens, it is expanded into a wave which appears to diverge from the focal point of the lens. Thus, when the recorded pattern is illuminated with the original plane wave, some of the light is diffracted into a diverging beam equivalent to the original plane wave; a holographic recording of the point source has been created.

When the plane wave is incident at a non-normal angle, the pattern formed is more complex but still acts as a negative lens provided it is illuminated at the original angle.



Sinusoidal zone plate

## Complex objects

To record a hologram of a complex object, a laser beam is first split into two separate beams of light. One beam illuminates the object, which then scatters light onto the recording medium. According to diffraction theory, each point in the object acts as a point source of light so the recording medium can be considered to be illuminated by a set of point sources located at varying distances from the medium.

The second (reference) beam illuminates the recording medium directly. Each point source wave interferes with the reference beam, giving rise to its own sinusoidal zone plate in the recording medium. The resulting pattern is the sum of all these 'zone plates' which combine to produce a random (speckle) pattern as in the photograph above.

When the hologram is illuminated by the original reference beam, each of the individual zone plates reconstructs the object wave which produced it, and these individual wavefronts add together to reconstruct the whole of the object beam. The viewer perceives a wavefront that is identical to the wavefront scattered from the object onto the recording medium, so that it appears to him or her that the object is still in place even if it has been removed. This image is known as a "virtual" image, as it is generated even though the object is no longer there.

## Mathematical model

A single-frequency light wave can be modelled by a complex number  $U$ , which represents the electric or magnetic field of the light wave. The amplitude and phase of the light are represented by the absolute value and angle of the complex number. The object and reference waves at any point in the holographic system are given by  $U_O$  and  $U_R$ . The combined beam is given by  $U_O + U_R$ . The energy of the combined beams is proportional to the square of magnitude of the combined waves as:

$$|U_O + U_R|^2 = U_O U_R^* + |U_R|^2 + |U_O|^2 + U_O^* U_R$$

If a photographic plate is exposed to the two beams and then developed, its transmittance,  $T$ , is proportional to the light energy that was incident on the plate and is given by

$$T = k U_O U_R^* + k |U_R|^2 + k |U_O|^2 + k U_O^* U_R$$

where  $k$  is a constant.

When the developed plate is illuminated by the reference beam, the light transmitted through the plate,  $U_H$  is equal to the transmittance  $T$  multiplied by the reference beam amplitude  $U_R$ , giving

$$U_H = T U_R = k U_O |U_R|^2 + k |U_R|^2 U_R + k |U_O|^2 U_R + k U_O^* U_R^2$$

It can be seen that  $U_H$  has four terms, each representing a light beam emerging from the hologram. The first of these is proportional to  $U_O$ . This is the reconstructed object beam which enables a viewer to 'see' the original object even when it is no longer present in the field of view.

The second and third beams are modified versions of the reference beam. The fourth term is known as the "conjugate object beam". It has the reverse curvature to the object beam itself and forms a real image of the object in the space beyond the holographic plate.

When the reference and object beams are incident on the holographic recording medium at significantly different angles, the virtual, real and reference wavefronts all emerge at different angles, enabling the reconstructed object to be seen clearly.

## Recording a hologram

### Items required

To make a hologram, the following are required:

- a suitable object or set of objects
- a suitable laser beam
- part of the laser beam to be directed so that it illuminates the object (the object beam) and another part so that it illuminates the recording medium directly (the reference beam), enabling the reference beam and the light which is scattered from the object onto the recording medium to form an interference pattern
- a recording medium which converts this interference pattern into an optical element which modifies either the amplitude or the phase of an incident light beam according to the intensity of the interference pattern.
- an environment which provides sufficient mechanical and thermal stability that the interference pattern is stable during the time in which the interference pattern is recorded<sup>[21]</sup>



An optical table being used to make a hologram

These requirements are inter-related, and it is essential to understand the nature of optical interference to see this. Interference is the variation in intensity which can occur when two light waves are superimposed. The intensity of the maxima exceeds the sum of the individual intensities of the two beams, and the intensity at the minima is less than this and may be zero. The interference pattern maps the relative phase between the two waves, and any change in the relative phases causes the interference pattern to move across the field of view. If the relative phase of the two waves changes by one cycle, then the pattern drifts by one whole fringe. One phase cycle corresponds to a change in the relative distances travelled by the two beams of one wavelength. Since the wavelength of light is of the order of  $0.5\mu\text{m}$ , it can be seen that very small changes in the optical paths travelled by either of the beams in the holographic recording system lead to movement of the interference pattern which is the holographic recording. Such changes can be caused by relative movements of any of the optical components or the object itself, and also by local changes in air-temperature. It is essential that any such changes are significantly less than the wavelength of light if a clear well-defined recording of the interference is to be created.

The exposure time required to record the hologram depends on the laser power available, on the particular medium used and on the size and nature of the object(s) to be recorded, just as in conventional photography. This determines the stability requirements. Exposure times of several minutes are typical when using quite powerful gas lasers and silver halide emulsions. All the elements within the optical system have to be stable to fractions of a  $\mu\text{m}$  over that period. It is possible to make holograms of much less stable objects by using a pulsed laser which produces a large amount of energy in a very short time ( $\mu\text{s}$  or less).<sup>[22]</sup> These systems have been used to produce holograms of live people. A holographic portrait of Dennis Gabor was produced in 1971 using a pulsed ruby laser.<sup>[23][24]</sup>

Thus, the laser power, recording medium sensitivity, recording time and mechanical and thermal stability requirements are all interlinked. Generally, the smaller the object, the more compact the optical layout, so that the stability requirements are significantly less than when making holograms of large objects.

Another very important laser parameter is its coherence.<sup>[25]</sup> This can be envisaged by considering a laser producing a sine wave whose frequency drifts over time; the coherence length can then be considered to be the distance over which it maintains a single frequency. This is important because two waves of different frequencies do not produce a stable interference pattern. The coherence length of the laser determines the depth of field which can be recorded in the scene. A good holography laser will typically have a coherence length of several meters, ample for a deep hologram.

The objects that form the scene must, in general, have optically rough surfaces so that they scatter light over a wide range of angles. A specularly reflecting (or shiny) surface reflects the light in only one direction at each point on its

surface, so in general, most of the light will not be incident on the recording medium. A hologram of a shiny object can be made by locating it very close to the recording plate.<sup>[26]</sup>

## **Hologram classifications**

There are three important properties of a hologram which are defined in this section. A given hologram will have one or other of each of these three properties, e.g. we can have an amplitude modulated thin transmission hologram, or a phase modulated, volume reflection hologram.

### **Amplitude and phase modulation holograms**

An amplitude modulation hologram is one where the amplitude of light diffracted by the hologram is proportional to the intensity of the recorded light. A straightforward example of this is photographic emulsion on a transparent substrate. The emulsion is exposed to the interference pattern, and is subsequently developed giving a transmittance which varies with the intensity of the pattern - the more light that fell on the plate at a given point, the darker the developed plate at that point.

A phase hologram is made by changing either the thickness or the refractive index of the material in proportion to the intensity of the holographic interference pattern. This is a phase grating and it can be shown that when such a plate is illuminated by the original reference beam, it reconstructs the original object wavefront. The efficiency (i.e. the fraction of the illuminated beam which is converted to reconstructed object beam) is greater for phase than for amplitude modulated holograms.

### **Thin holograms and thick (volume) holograms**

A thin hologram is one where the thickness of the recording medium is much less than the spacing of the interference fringes which make up the holographic recording.

A thick or volume hologram is one where the thickness of the recording medium is greater than the spacing of the interference pattern. The recorded hologram is now a three dimensional structure, and it can be shown that incident light is diffracted by the grating only at a particular angle, known as the Bragg angle.<sup>[27]</sup> If the hologram is illuminated with a light source incident at the original reference beam angle but a broad spectrum of wavelengths, reconstruction occurs only at the wavelength of the original laser used. If the angle of illumination is changed, reconstruction will occur at a different wavelength and the colour of the re-constructed scene changes. A volume hologram effectively acts as a colour filter.

### **Transmission and reflection holograms**

A transmission hologram is one where the object and reference beams are incident on the recording medium from the same side. In practice, several more mirrors may be used to direct the beams in the required directions.

Normally, transmission holograms can only be reconstructed using a laser or a quasi-monochromatic source, but a particular type of transmission hologram, known as a rainbow hologram, can be viewed with white light.

In a reflection hologram, the object and reference beams are incident on the plate from opposite sides of the plate. The reconstructed object is then viewed from the same side of the plate as that at which the re-constructing beam is incident.

Only volume holograms can be used to make reflection holograms, as only a very low intensity diffracted beam would be reflected by a thin hologram.



## Holographic recording media

The recording medium has to convert the original interference pattern into an optical element that modifies either the amplitude or the phase of an incident light beam in proportion to the intensity of the original light field.

The recording medium should be able to resolve fully all the fringes arising from interference between object and reference beam. These fringe spacings can range from tens of microns to less than one micron, i.e. spatial frequencies ranging from a few hundred to several thousand cycles/mm, and ideally, the recording medium should have a response which is flat over this range. If the response of the medium to these spatial frequencies is low, the diffraction efficiency of the hologram will be poor, and a dim image will be obtained. It should be noted that standard photographic film has a very low, or even zero, response at the frequencies involved and cannot be used to make a hologram - see, for example, Kodak's professional black and white film<sup>[28]</sup> whose resolution starts falling off at 20 lines/mm — it is unlikely that any reconstructed beam could be obtained using this film.

If the response is not flat over the range of spatial frequencies in the interference pattern, then the resolution of the reconstructed image may also be degraded.<sup>[29][30]</sup>

The table below shows the principal materials used for holographic recording. Note that these do not include the materials used in the mass replication of an existing hologram, which are discussed in the next section. The resolution limit given in the table indicates the maximal number of interference lines/mm of the gratings. The required exposure, expressed as millijoules (mJ) of photon energy impacting the surface area, is for a long exposure time. Short exposure times (less than 1/1000 of a second, such as with a pulsed laser) require much higher exposure energies, due to reciprocity failure.

### General properties of recording materials for holography. Source:<sup>[31]</sup>

Material	Reusable	Processing	Type of hologram	Theoretical maximum efficiency	Required exposure [mJ/cm <sup>2</sup> ]	Resolution limit [mm <sup>-1</sup> ]
Photographic emulsions	No	Wet	Amplitude	6%	1.5	5000
			Phase (bleached)	60%		
Dichromated gelatin	No	Wet	Phase	100%	100	10,000
Photoresists	No	Wet	Phase	30%	100	3,000
Photothermoplastics	Yes	Charge and heat	Phase	33%	0.1	500–1,200
Photopolymers	No	Post exposure	Phase	100%	10000	5,000
Photorefractives	Yes	None	Phase	100%	10	10,000

## Copying and mass production

An existing hologram can be copied by embossing<sup>[32]</sup> or optically.<sup>[33]</sup>

Most holographic recordings (e.g. bleached silver halide, photoresist, and photopolymers) have surface relief patterns which conform with the original illumination intensity. Embossing, which is similar to the method used to stamp out plastic discs from a master in audio recording, involves copying this surface relief pattern by impressing it onto another material.

The first step in the embossing process is to make a stamper by electrodeposition of nickel on the relief image recorded on the photoresist or photothermoplastic. When the nickel layer is thick enough, it is separated from the master hologram and mounted on a metal backing plate. The material used to make embossed copies consists of a polyester base film, a resin separation layer and a thermoplastic film constituting the holographic layer.

The embossing process can be carried out with a simple heated press. The bottom layer of the duplicating film (the thermoplastic layer) is heated above its softening point and pressed against the stamper, so that it takes up its shape. This shape is retained when the film is cooled and removed from the press. In order to permit the viewing of embossed holograms in reflection, an additional reflecting layer of aluminum is usually added on the hologram recording layer. This method is particularly suited to mass production.

The first book to feature a hologram on the front cover was *The Skook* (Warner Books, 1984) by JP Miller, featuring an illustration by Miller. That same year, "Telstar" by Ad Infinitum became the first record with a hologram cover and *National Geographic* published the first magazine with a hologram cover.<sup>[34]</sup> Embossed holograms are used widely on credit cards, banknotes, and high value products for authentication purposes.<sup>[35]</sup>

It is possible to print holograms directly into steel using a sheet explosive charge to create the required surface relief.<sup>[36]</sup> The Royal Canadian Mint produces holographic gold and silver coinage through a complex stamping process.<sup>[37]</sup>

A hologram can be copied optically by illuminating it with a laser beam, and locating a second hologram plate so that it is illuminated both by the reconstructed object beam, and the illuminating beam. Stability and coherence requirements are significantly reduced if the two plates are located very close together.<sup>[38]</sup> An index matching fluid is often used between the plates to minimize spurious interference between the plates. Uniform illumination can be obtained by scanning point-by-point or with a beam shaped into a thin line.

## Reconstructing and viewing the holographic image

When the hologram plate is illuminated by a laser beam identical to the reference beam which was used to record the hologram, an exact reconstruction of the original object wavefront is obtained. An imaging system (an eye or a camera) located in the reconstructed beam 'sees' exactly the same scene as it would have done when viewing the original. When the lens is moved, the image changes in the same way as it would have done when the object was in place. If several objects were present when the hologram was recorded, the reconstructed objects move relative to one another, i.e. exhibit parallax, in the same way as the original objects would have done. It was very common in the early days of holography to use a chess board as the object and then take photographs at several different angles using the reconstructed light to show how the relative positions of the chess pieces appeared to change.

A holographic image can also be obtained using a different laser beam configuration to the original recording object beam, but the reconstructed image will not match the original exactly.<sup>[39]</sup> When a laser is used to reconstruct the hologram, the image is speckled just as the original image will have been. This can be a major drawback in viewing a hologram.

White light consists of light of a wide range of wavelengths. Normally, if a hologram is illuminated by a white light source, each wavelength can be considered to generate its own holographic reconstruction, and these will vary in size, angle, and distance. These will be superimposed, and the summed image will wipe out any information about the original scene, just as if you superimposed a set of photographs of the same object of different sizes and orientations. However, a holographic image can be obtained using white light in specific circumstances, e.g. with volume holograms and rainbow holograms. The white light source used to view these holograms should always approximate to a point source, i.e. a spot light or the sun. An extended source (e.g. a fluorescent lamp) will not reconstruct a hologram since its light is incident at each point at a wide range of angles, giving multiple reconstructions which will "wipe" one another out.

White light reconstructions do not contain speckles.

## Volume holograms

A volume hologram can give a reconstructed beam using white light, as the hologram structure effectively filters out colours other than those equal to or very close to the colour of the laser used to make the hologram so that the reconstructed image will appear to be approximately the same colour as the laser light used to create the holographic recording.

## Rainbow holograms

In this method, parallax in the vertical plane is sacrificed to allow a bright well-defined single colour re-constructed image to be obtained using white light. The rainbow holography recording process uses a horizontal slit to eliminate vertical parallax in the output image. The viewer is then effectively viewing the holographic image through a narrow horizontal slit. Horizontal parallax information is preserved but movement in the vertical direction produces colour rather than different vertical perspectives.<sup>[40]</sup> Stereopsis and horizontal motion parallax, two relatively powerful cues to depth, are preserved.

The holograms found on credit cards are examples of rainbow holograms. These are technically transmission holograms mounted onto a reflective surface like a metalized polyethylene terephthalate substrate commonly known as PET.



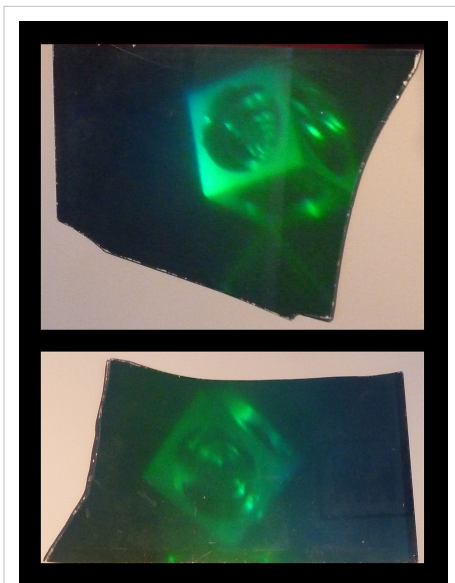
Rainbow hologram showing the change in colour in the vertical direction

## Fidelity of the reconstructed beam

To replicate the original object beam exactly, the reconstructing reference beam must be identical to the original reference beam and the recording medium must be able to fully resolve the interference pattern formed between the object and reference beams. Exact reconstruction is required in holographic interferometry, where the holographically reconstructed wavefront interferes with the wavefront coming from the actual object, giving a null fringe if there has been no movement of the object and mapping out the displacement if the object has moved. This requires very precise relocation of the developed holographic plate.

Any change in the shape, orientation or wavelength of the reference beam gives rise to aberrations in the reconstructed image. For instance, the reconstructed image is magnified if the laser used to reconstruct the hologram has a shorter wavelength than the original laser. Nonetheless, good reconstruction is obtained using a laser of a different wavelength, quasi-monochromatic light or white light, in the right circumstances.

Since each point in the object illuminates all of the hologram, the whole object can be reconstructed from a small part of the hologram. Thus, a hologram can be broken up into small pieces and each one will enable the whole of the original object to be imaged. One does, however, lose information and the spatial resolution gets worse as the size of the hologram is decreased — the image becomes "fuzzier". The field of view is also reduced, and the viewer will have to change position to see different parts of the scene.



Reconstructions from two parts of a broken hologram. Note the different viewpoints required to see the whole object

## Applications

### Art

Early on, artists saw the potential of holography as a medium and gained access to science laboratories to create their work. Holographic art is often the result of collaborations between scientists and artists, although some holographers would regard themselves as both an artist and a scientist.

Salvador Dalí claimed to have been the first to employ holography artistically. He was certainly the first and best-known surrealist to do so, but the 1972 New York exhibit of Dalí holograms had been preceded by the holographic art exhibition that was held at the Cranbrook Academy of Art in Michigan in 1968 and by the one at the Finch College gallery in New York in 1970, which attracted national media attention.<sup>[41]</sup>

During the 1970s, a number of art studios and schools were established, each with their particular approach to holography. Notably, there was the San Francisco School of Holography established by Lloyd Cross, The Museum of Holography in New York founded by Rosemary (Possie) H. Jackson, the Royal College of Art in London and the Lake Forest College Symposiums organised by Tung Jeong (T.J.).<sup>[42]</sup> None of these studios still exist; however, there is the Center for the Holographic Arts in New York<sup>[43]</sup> and the HOLOcenter in Seoul,<sup>[44]</sup> which offers artists a place to create and exhibit work.

During the 1980s, many artists who worked with holography helped the diffusion of this so-called "new medium" in the art world, such as Harriet Casdin-Silver of the USA, Dieter Jung of Germany, and Moysés Baumstein of Brazil, each one searching for a proper "language" to use with the three-dimensional work, avoiding the simple holographic reproduction of a sculpture or object. For instance, in Brazil, many concrete poets (Augusto de Campos, Décio Pignatari, Julio Plaza and José Wagner Garcia, associated with Moysés Baumstein) found in holography a way to express themselves and to renew Concrete Poetry.

A small but active group of artists still use holography as their main medium, and many more artists integrate holographic elements into their work.<sup>[45]</sup> Some are associated with novel holographic techniques; for example, artist Matt Brand<sup>[46]</sup> employed computational mirror design to eliminate image distortion from specular holography.

The MIT Museum<sup>[47]</sup> and Jonathan Ross<sup>[48]</sup> both have extensive collections of holography and on-line catalogues of art holograms.

### Data storage

Holography can be put to a variety of uses other than recording images. Holographic data storage is a technique that can store information at high density inside crystals or photopolymers. The ability to store large amounts of information in some kind of media is of great importance, as many electronic products incorporate storage devices. As current storage techniques such as Blu-ray Disc reach the limit of possible data density (due to the diffraction-limited size of the writing beams), holographic storage has the potential to become the next generation of popular storage media. The advantage of this type of data storage is that the volume of the recording media is used instead of just the surface. Currently available SLMs can produce about 1000 different images a second at 1024×1024-bit resolution. With the right type of media (probably polymers rather than something like LiNbO<sub>3</sub>), this would result in about one-gigabit-per-second writing speed. Read speeds can surpass this, and experts believe one-terabit-per-second readout is possible. In 2005, companies such as Optware and Maxell produced a 120 mm disc that uses a holographic layer to store data to a potential 3.9 TB, which they plan to market under the name Holographic Versatile Disc. Another company, InPhase Technologies, is developing a competing format. While many holographic data storage models have used "page-based" storage, where each recorded hologram holds a large amount of data, more recent research into using submicrometre-sized "microholograms" has resulted in several potential 3D optical data storage solutions. While this approach to data storage can not attain the high data rates of page-based storage, the tolerances, technological hurdles, and cost of producing a commercial product are significantly lower.

## Dynamic holography

In static holography, recording, developing and reconstructing occur sequentially, and a permanent hologram is produced.

There also exist holographic materials that do not need the developing process and can record a hologram in a very short time. This allows one to use holography to perform some simple operations in an all-optical way. Examples of applications of such real-time holograms include phase-conjugate mirrors ("time-reversal" of light), optical cache memories, image processing (pattern recognition of time-varying images), and optical computing.

The amount of processed information can be very high (terabits/s), since the operation is performed in parallel on a whole image. This compensates for the fact that the recording time, which is in the order of a microsecond, is still very long compared to the processing time of an electronic computer. The optical processing performed by a dynamic hologram is also much less flexible than electronic processing. On one side, one has to perform the operation always on the whole image, and on the other side, the operation a hologram can perform is basically either a multiplication or a phase conjugation. In optics, addition and Fourier transform are already easily performed in linear materials, the latter simply by a lens. This enables some applications, such as a device that compares images in an optical way.<sup>[49]</sup>

The search for novel nonlinear optical materials for dynamic holography is an active area of research. The most common materials are photorefractive crystals, but in semiconductors or semiconductor heterostructures (such as quantum wells), atomic vapors and gases, plasmas and even liquids, it was possible to generate holograms.

A particularly promising application is optical phase conjugation. It allows the removal of the wavefront distortions a light beam receives when passing through an aberrating medium, by sending it back through the same aberrating medium with a conjugated phase. This is useful, for example, in free-space optical communications to compensate for atmospheric turbulence (the phenomenon that gives rise to the twinkling of starlight).

## Hobbyist use

Since the beginning of holography, experimenters have explored its uses. Starting in 1971, Lloyd Cross started the San Francisco School of Holography and started to teach amateurs the methods of making holograms with inexpensive equipment. This method relied on the use of a large table of deep sand to hold the optics rigid and damp vibrations that would destroy the image.

Many of these holographers would go on to produce art holograms. In 1983, Fred Unterseher published the *Holography Handbook*, a remarkably easy-to-read description of making holograms at home. This brought in a new wave of holographers and gave simple methods to use the then-available AGFA silver halide recording materials.

In 2000, Frank DeFreitas published the *Shoebox Holography Book* and introduced the use of inexpensive laser pointers to countless hobbyists.

This was a very important development for amateurs, as the cost for a 5 mW laser dropped from \$1200 to \$5 as semiconductor laser diodes reached mass market. Now, there are hundreds to thousands of amateur holographers worldwide.

In 2006, a large number of surplus Holography Quality Green Lasers (Coherent C315) became available and put Dichromated Gelatin (DCG) within the reach of the amateur holographer. The holography community was surprised at the amazing sensitivity of DCG to green light. It had been assumed that the sensitivity would be non-existent. Jeff Blyth responded with the G307 formulation of DCG to increase the speed and sensitivity to these new lasers.<sup>[50]</sup>



*Peace Within Reach*, a Denisjuk DCG hologram  
by amateur Dave Battin

Many film suppliers have come and gone from the silver-halide market. While more film manufactures have filled in the voids, many amateurs are now making their own film. The favorite formulations are Dichromated Gelatin, Methylene Blue Sensitised Dichromated Gelatin and Diffusion Method Silver Halide preparations. Jeff Blyth has published very accurate methods for making film in a small lab or garage.<sup>[51]</sup>

A small group of amateurs are even constructing their own pulsed lasers to make holograms of moving objects.<sup>[52]</sup>

Holography kits with self-developing film plates have now entered the consumer market. The kits make holographs and have been found to be fairly error tolerant,<sup>[53]</sup> and enable holograms to be made without any other specialized equipment.

### Holographic interferometry

Holographic interferometry (HI) is a technique that enables static and dynamic displacements of objects with optically rough surfaces to be measured to optical interferometric precision (i.e. to fractions of a wavelength of light).<sup>[54][55]</sup> It can also be used to detect optical-path-length variations in transparent media, which enables, for example, fluid flow to be visualized and analyzed. It can also be used to generate contours representing the form of the surface.

It has been widely used to measure stress, strain, and vibration in engineering structures.

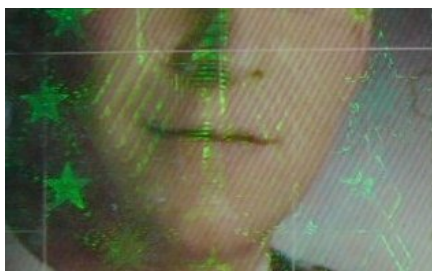
### Interferometric microscopy

The hologram keeps the information on the amplitude and phase of the field. Several holograms may keep information about the same distribution of light, emitted to various directions. The numerical analysis of such holograms allows one to emulate large numerical aperture, which, in turn, enables enhancement of the resolution of optical microscopy. The corresponding technique is called interferometric microscopy. Recent achievements of interferometric microscopy allow one to approach the quarter-wavelength limit of resolution.<sup>[56]</sup>

### Sensors or biosensors

The hologram is made with a modified material that interacts with certain molecules generating a change in the fringe periodicity or refractive index, therefore, the color of the holographic reflection.<sup>[57]</sup>

### Security



*Identigram* as a security element in a German identity card

Security holograms are very difficult to forge, because they are replicated from a master hologram that requires expensive, specialized and technologically advanced equipment. They are used widely in many currencies, such as the Brazilian 20, 50, and 100-reais notes; British 5, 10, and 20-pound notes; South Korean 5000, 10000, and 50000-won notes; Japanese 5000 and 10000 yen notes; and all the currently-circulating banknotes of the Canadian dollar, Danish krone, and Euro. They can also be found in credit and bank cards as well as passports, ID cards, books, DVDs, and sports equipment.

## Other applications

Holographic scanners are in use in post offices, larger shipping firms, and automated conveyor systems to determine the three-dimensional size of a package. They are often used in tandem with checkweighers to allow automated pre-packing of given volumes, such as a truck or pallet for bulk shipment of goods. Holograms produced in elastomers can be used as stress-strain reporters due to its elasticity and compressibility, the pressure and force applied are correlated to the reflected wavelength, therefore its color.<sup>[58]</sup>

## Non-optical holography

In principle, it is possible to make a hologram for any wave.

Electron holography is the application of holography techniques to electron waves rather than light waves. Electron holography was invented by Dennis Gabor to improve the resolution and avoid the aberrations of the transmission electron microscope. Today it is commonly used to study electric and magnetic fields in thin films, as magnetic and electric fields can shift the phase of the interfering wave passing through the sample.<sup>[59]</sup> The principle of electron holography can also be applied to interference lithography.<sup>[60]</sup>

Acoustic holography is a method used to estimate the sound field near a source by measuring acoustic parameters away from the source via an array of pressure and/or particle velocity transducers. Measuring techniques included within acoustic holography are becoming increasingly popular in various fields, most notably those of transportation, vehicle and aircraft design, and NVH. The general idea of acoustic holography has led to different versions such as near-field acoustic holography (NAH) and statistically optimal near-field acoustic holography (SONAH). For audio rendition, the wave field synthesis is the most related procedure.

*Atomic holography* has evolved out of the development of the basic elements of atom optics. With the Fresnel diffraction lens and atomic mirrors atomic holography follows a natural step in the development of the physics (and applications) of atomic beams. Recent developments including atomic mirrors and especially ridged mirrors have provided the tools necessary for the creation of atomic holograms,<sup>[61]</sup> although such holograms have not yet been commercialized.

## Things often confused with holograms

Effects produced by lenticular printing, the Pepper's Ghost illusion (or modern variants such as the Musion Eyeliner), tomography and volumetric displays are often confused with holograms.<sup>[62][63]</sup>

The Pepper's ghost technique, being the easiest to implement of these methods, is most prevalent in 3D displays that claim to be (or are referred to as) "holographic". While the original illusion, used in theater, recurred to actual physical objects and persons, located offstage, modern variants replace the source object with a digital screen, which displays imagery generated with 3D computer graphics to provide the necessary depth cues. The reflection, which seems to float mid-air, is still flat, however, thus less realistic than if an actual 3D object was being reflected.

Examples of this digital version of Pepper's ghost illusion include the Gorillaz performances in the 2005 MTV Europe Music Awards and the 48th Grammy Awards; and Tupac Shakur's virtual performance at Coachella Valley Music and Arts Festival in 2012, rapping alongside Snoop Dogg during the latter's set with Dr. Dre.<sup>[64]</sup>

During the 2008 American presidential election, CNN debuted its tomograms to "beam in" correspondents including musician will.i.am as "holograms".

An even simpler illusion can be created by rear-projecting realistic images into semi-transparent screens. The rear projection is necessary because otherwise the semi-transparency of the screen would allow the background to be illuminated by the projection, which would break the illusion.

Crypton Future Media, a music software company that produced Hatsune Miku,<sup>[65]</sup> one of many Vocaloid singing synthesizer applications, has produced concerts that have Miku, along with other Crypton Vocaloids, performing on

stage as "holographic" characters. These concerts use rear projection onto a semi-transparent DILAD screen<sup>[66][67]</sup> to achieve its "holographic" effect.<sup>[68][69][70]</sup>

In 2011, in Beijing, apparel company Burberry produced the "Burberry Prorsum Autumn/Winter 2011 Hologram Runway Show", which included life size 2-D projections of models. The company's own video<sup>[71]</sup> shows several centered and off-center shots of the main 2-dimensional projection screen, the latter revealing the flatness of the virtual models. The claim that holography was used was reported as fact in the trade media.<sup>[72]</sup>

## Holography in fiction

Holograms are often used as plot devices in science fiction. However, very often, sci-fi movies and TV shows incorrectly present different 3D-projection technologies as "holography" (see the above section).

- *The Carpathian Castle* (1893 novel by *Jules Verne*), the plot revolves around prima donna La Stilla, represented at the times of the events as a projected image.
- *The Jetsons* (1962-3 television series), holograms used as entertainment devices, replacing the television in many episodes
- *Star Trek: The Animated Series* (1974 television series) episode "The Practical Joker", the holodeck is introduced
- *Star Wars* (1977 film), use of the hologram in the movies and video games of the series to display people remotely communicating with each another. It is also present in several other of the films of the series, including the Phantom Menace, Attack of the Clones, and Revenge of the Sith to communicate across the galaxy.
- *Hello America* (1981 book by *J.G. Ballard*), holographic technology is used by president *Charles Manson* to scare nomad peoples along the *United States of America*, showing images of American pop culture icons such as Gary Cooper, Mickey Mouse, or the Enterprise space ship.
- *Jem and the Holograms* (1985 television series), Jerrica Benton, the lead singer of a band uses hologram projections to help create her alter-ego persona, Jem; micro-projectors in her earrings allow her to project a hologram over herself and produce hologram objects and images in her surroundings
- *Star Trek: The Next Generation* (1987-1994 television series), uses the holodeck extensively; beginning with this series, various episodes and films throughout the *Star Trek* series feature holographic characters and ships
- *Red Dwarf* (1988 television series), after a catastrophic radiation leak inside the Jupiter Mining Corporation spaceship *Red Dwarf*, crew member Second Technician Arnold Rimmer is resurrected as a hologram. Because he is a "soft-light" hologram, he cannot touch anything and objects just pass right through him. However, much later - in series VI - the Red Dwarf crew meet 'Legion', a being with advanced technology, who upgrades Rimmer's *light bee* – the small object that projects his hologram by hovering around inside him – changing his projection to what is called in the show "hard-light" giving him a hologramatic equivalent of a physical body.
- *Back to the Future Part II* (1989 film), a giant projection hologram is used as an advertisement for the (fictional) 2015 film *Jaws 19*
- *Total Recall* (1990 film), the main character uses a device, similar to a wrist watch, to produce a hologram of himself and deceive his foes
- *Star Trek: Voyager* (1995–2001 television series), introduced the Emergency Medical Hologram (EMH) doctor
- *Yu-Gi-Oh!* (1996–present manga, film, television series, video games), use of holographic technology used in order to make a game called *Duel Monsters* appear to be more life like; *Duel Monsters* is a game where players using a wrist mounted Duel Disk summon monsters and cast spells and traps in order to bring a players life points to 0 or diminish all the cards in a players deck; used throughout the entire series.
- *Stargate: SG-1* (1997–2007 television series), various characters appear as holograms in various episodes: The Asgard masquerade themselves holographically as Norse gods to the primitive peoples under their protection, Morgan le Fay in "The Pegasus Project" and Myrddin as a Merlin in "Avalon" and "Camelot" as a holographic sentry; Heliopolis "Book"; the puddle jumper starship has a holographic HUD. After the Goa'uld leader Anubis probed the mind of Asgard leader Thor, he was able to acquire their hologram technology and he used it



frequently.

- *Half-Life* (1998 video game), the scientific research company Black Mesa is known to use holograms as recorded messages in their facility
  - *Lost in Space* (1998 film), June Lockhart (Maureen Robinson) appeared as Will's school principal "Cartwright" in a hologram
  - *Power Rangers Time Force* (2001 television series), their chrono morphers use holographic communication
  - *Halo* (2001 video game) uses "holotanks" to display the avatar of an artificial intelligence construct called Cortana; in *Halo: Reach*, an armor ability called the hologram allows the user to create an identical decoy
  - *Vanilla Sky* (2001 film), a holographic projection of jazz musician John Coltrane appears in the main character's apartment during his birthday party
  - *The First \$20 Million Is Always the Hardest* (2002 film), computer geeks develop a \$99 computer using a holographic projector as both the display and user interface
  - *Treasure Planet* (2002 film), Jim as a little boy reads from a 3D hologram book the story about Captain Flint and Treasure Planet; later, Jim as a teenager finds a sphere map and uses it to look at the galaxy map to Treasure Planet
  - *Pinocchio 3000* (2004 film), Mayor Scamboli owns a 3D hologram map on his table; Cabby and Roto change channels on it; later, at Scamboland, Mayor Scamboli welcomes kids as a giant 3D hologram version for Scamboland carnival opening
  - *Stargate: Atlantis* (2004–2009 television series), the Atlantis city-starship features a hologram room that allows access to the Ancient database in the form of holograms; an Ancient Control Chair contains holographic projectors; in the episode "Rising", Melia (a member of the Atlantean High Council during the first siege of Atlantis some ten millennia ago) is first seen as a hologram describing the history of the Ancients in the Pegasus Galaxy; Aurora-class battleship can project holograms remotely for communication purposes
  - *The Island* (2005 film), a holographic projector surrounded the military compound where clones were kept to give the illusion of a tropical environment; holographic displays are present on various terminals, including the MSN information terminal in Los Angeles
  - *Meet the Robinsons* (2007 film), Bowler Hat Michael Goob Yagoobian has a discussion with the Bowler Hat Robot about getting revenge and Bowler Hat robot shows him a 3D hologram image of a flying car-plane time machine
  - *Mass Effect* series (2007-12 video game series), computer GUIs are explained in the codex to consist of a projected holographic display, combined with the use of force feedback gloves that allow the user to experience simulated tactile sensations when manipulating them; the game's "omni-tools" are holographic user-interfaces that act as a cell-phone-like device, serving as a port between computers, tablets, and other devices, giving the user many capabilities, including the ability to transfer information via wireless technology to others with an omni-tool
  - *Dead Space* (2008 video game), to replace the player's HUD, a holographic display shows up in front of the player's character
  - *Iron Man* (2008 film) and *Iron Man 2* (2010 film), holographic displays appear in Iron Man's suit
  - *Avatar* (2009 film), holographic displays are used extensively on terminals and HUDs
  - *G.I. Joe: The Rise of Cobra* (2009 film), Hawk, Destro, Baroness, and Storm Shadow appear as holographic projections
  - *Enthiran* (2010 film), Chitti the robot can be telecommunicated with using a "virtual calling" where each caller can be seen as a holographic projection in front of the robot during the call
-

## References

- [1] Gabor, Dennis. (1948), A new microscopic principle, *Nature*, 161, p 777-8
- [2] Gabor, Dennis (1949). "Microscopy by reconstructed wavefronts". *Proceedings of the Royal Society (London)* **197** (1051): 454–487. Bibcode 1949RSPSA.197..454G. doi:10.1098/rspa.1949.0075
- [3] "The Nobel Prize in Physics 1971" ([http://www.nobelprize.org/nobel\\_prizes/physics/laureates/1971/](http://www.nobelprize.org/nobel_prizes/physics/laureates/1971/)). Nobelprize.org. . Retrieved 2012-04-21.
- [4] Hariharan, (1996), Section 1.2, p4-5
- [5] Denisuyuk, Yuri N. (1962). "On the reflection of optical properties of an object in a wave field of light scattered by it". *Doklady Akademii Nauk SSSR* **144** (6): 1275–1278.
- [6] Leith, E.N.; Upatnieks, J. (1962). "Reconstructed wavefronts and communication theory". *J. Opt. Soc. Am.* **52** (10): 1123–1130. doi:10.1364/JOSA.52.001123.
- [7] Upatniek J & Leonard C., (1969), "Diffraction efficiency of bleached photographically recorded interference patterns", *Applied Optics*, 8, p85-89
- [8] Graube A, (1974), "Advances in bleaching methods for photographically recorded holograms", *Applied Optics*, 13, p2942-6
- [9] N. J. Phillips and D. Porter, (1976), "An advance in the processing of holograms," *Journal of Physics E: Scientific Instruments* p. 631
- [10] Hariharan, (2002), Section 7.1, p 60
- [11] Benton S.A, (1977), "White light transmission/reflection holography" in *Applications of Holography and Optical Data Processing*, ed. E. Marom et al, ps 401-9, Pergamon Press, Oxford
- [12] Toal Vincent (2012), "Introduction to Holography", CRC Press, ISBN 978-1-4398-1868-8
- [13] Hariharan, (2002), Section 7.2, p61
- [14] "specular holography: how" (<http://www.zintaglio.com/how.html>). Zintaglio.com. . Retrieved 2012-04-21.
- [15] "MIT unveils holographic TV system" ([http://www.tgdaily.com/hardware-features/53703-mit-unveils-holographic-tv-system?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed:+tgdaily\\_all\\_sections+\(TG+Daily+-+All+News\)\)](http://www.tgdaily.com/hardware-features/53703-mit-unveils-holographic-tv-system?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+tgdaily_all_sections+(TG+Daily+-+All+News)))). . Retrieved 2011-09-14.
- [16] See Zebra imaging.
- [17] Blanche, P.-A.; Bablumian, A.; Voorakaranam, R.; Christenson, C.; Lin, W.; Gu, T.; Flores, D.; Wang, P. et al. (2010). "Holographic three-dimensional telepresence using large-area photorefractive polymer" (<http://www.nature.com/nature/journal/v468/n7320/full/nature09521.html>). *Nature* **468** (7320): 80–83. Bibcode 2010Natur.468...80B. doi:10.1038/nature09521. PMID 21048763. .
- [18] Hariharan, (2002), 12.6, p107
- [19] "Holography" (<http://hyperphysics.phy-astr.gsu.edu/Hbase/optmod/holog.html>). Hyperphysics.phy-astr.gsu.edu. . Retrieved 2012-04-21.
- [20] Hariharan, (2002), Section 1, p1
- [21] Hariharan, (2002), Section 7.1, p60
- [22] Martinez-Hurtado et al. doi:10.1021/la102693m
- [23] Hariharan, (2002), Figure 4.5, p44
- [24] "Photograph of Dennis Gabor standing beside his holographic portrait" (<http://webmuseum.mit.edu/browser.php?m=objects&kv=67243&i=14558>). MIT. . Retrieved 2011-09-16.
- [25] Hariharan, (2002), Section 4.2, p40
- [26] Hariharan, (2002), Figure 7.2, p62
- [27] Lipson, (2011), Section 12.5.4, p443
- [28] "Kodak black and white professional film" (<http://www.kodak.com/eknec/documents/59/0900688a80300559/EpubBW400CN4036.pdf>). . Retrieved 2011-09-14.
- [29] Hariharan, (1996), Section 6.4, p88
- [30] Kozma A & Zelenka JS, (1970), Effect of film resolution and size in holography, *Journal of the Optical Society of America*, 60, 34–43
- [31] Hariharan, (2002), Table 6.1, p50
- [32] Iwata F & Tsujiiuchi J (1974), "Characteristics of a photoresist hologram and its replica", *Applied Optics*, 13, p1327-36
- [33] Hariharan, (2002), Section 11.4.1, p191
- [34] Freitas, Frank De (2008-07-30). "Antiquarian Holographica blog" (<http://holographica.blogspot.com/2008/07/national-geographic-hologram-1984.html>). Holographica.blogspot.com. . Retrieved 2012-04-21.
- [35] Toal Vincent, 2012, Introduction to Holography, CRC Press, ISBN 978-1-4398-1868-8
- [36] "Holograms with explosive power" (<http://www.physorg.com/news124039000.html>). Physorg.com. . Retrieved 2012-04-21.
- [37] "Lunar Holographic Coins" (<http://www.mint.ca/store/coin/150-lunar-hologram-coin-year-of-the-rabbit-2011-prod990012>). . Retrieved 2011-09-14.
- [38] Harris JR, Sherman GC and Billings BH, 1966, Copying hologram, *Applied Optics*, 5, 665-6
- [39] Hariharan, (2002), Section 2.3, p17
- [40] Hariharan, (2002), Section 7.4, p63
- [41] "The History and Development of Holography" (<http://holophile.com/history.htm>). Holophile.com. . Retrieved 2012-04-21.
- [42] Integraf. "Dr. Tung H. Jeong Biography" ([http://www.integraf.com/tung\\_jeong.htm](http://www.integraf.com/tung_jeong.htm)). Integraf.com. . Retrieved 2012-04-21.

- [43] "holocenter" (<http://www.holocenter.org>). holocenter. . Retrieved 2012-04-21.
- [44] "Holocenter" (<http://www.holocenter.or.kr/>). Holocenter. . Retrieved 2012-04-21.
- [45] <http://www.universal-hologram.com/>
- [46] Holographic metalwork <http://www.zintaglio.com>
- [47] "MIT Museum: Collections - Holography" (<http://web.mit.edu/museum/collections/holography.html>). Web.mit.edu. . Retrieved 2012-04-21.
- [48] "The Jonathan Ross Hologram Collection" (<http://www.jrholocollection.com/>). Jrholocollection.com. . Retrieved 2012-04-21.
- [49] R. Ryf et al. High-frame-rate joint Fourier-transform correlator based on  $\text{Sn}_2\text{P}_2\text{S}_6$  crystal (<http://ol.osa.org/abstract.cfm?id=65530>), *Optics Letters* **26**, 1666–1668 (2001)
- [50] "A Holography FAQ" ([http://www.holowiki.com/index.php/G307\\_DCG\\_Formula](http://www.holowiki.com/index.php/G307_DCG_Formula)). HoloWiki. 2011-02-15. . Retrieved 2012-04-21.
- [51] "Many methods are here" (<http://www.holowiki.com/index.php/Special:Search?search=Blyth&go=Go>). Holowiki.com. . Retrieved 2012-04-21.
- [52] "Jeff Blyth's Film Formulations" (<http://cabd0.tripod.com/holograms/index.html>). Cabd0.tripod.com. . Retrieved 2012-04-21.
- [53] <http://www.litholo.com/Hologram%20Kit%20article%20Physics%20Teacher%20Nov%202010.pdf>
- [54] Powell RL & Stetson KA, 1965, *J. Opt. Soc. Am.*, **55**, 1593–8
- [55] Jones R and Wykes C, *Holographic and Speckle Interferometry*, 1989, Cambridge University Press ISBN 0-521-34417-4
- [56] Y.Kuznetsova; A.Neumann, S.R.Brueck (2007). "Imaging interferometric microscopy—approaching the linear systems limits of [[optical resolution (<http://www.opticsexpress.org/abstract.cfm?id=134719>)]". *Optics Express* **15** (11): 6651–6663. Bibcode 2007OExpr..15.6651K. doi:10.1364/OE.15.006651. PMID 19546975. .
- [57] Martinez-Hurtado et al 2010; <http://pubs.acs.org/doi/abs/10.1021/la102693m>
- [58] 'Elastic hologram' pages 113–117, Proc. of the IGC 2010, ISBN 978-0-9566139-1-2 here: <http://www.dspace.cam.ac.uk/handle/1810/225960>
- [59] R. E. Dunin-Borkowski et al., *Micros. Res. and Tech.* vol. 64, pp. 390–402 (2004)
- [60] K. Ogai et al., *Jpn. J. Appl. Phys.*, vol. 32, pp.5988–5992 (1993)
- [61] F. Shimizu; J.Fujita (March 2002). "Reflection-Type Hologram for Atoms". *Physical Review Letters* **88** (12): 123201. Bibcode 2002PhRvL..8813201S. doi:10.1103/PhysRevLett.88.123201. PMID 11909457.
- [62] "Holographic announcers at Luton airport" (<http://www.bbc.co.uk/news/business-12328160>). Bbc.co.uk. 2011-01-31. . Retrieved 2012-04-21.
- [63] Farivar, Cyrus (2012-04-16). "Tupac "hologram" merely pretty cool optical illusion" (<http://arstechnica.com/science/news/2012/04/tupac-hologram-merely-pretty-cool-optical-illusion.ars>). Arstechnica.com. . Retrieved 2012-04-21.
- [64] "Tupac returns as a hologram at Coachella" (<http://marquee.blogs.cnn.com/2012/04/16/tupac-returns-as-a-hologram-at-coachella/>). *The Marquee Blog - CNN.com Blogs* (CNN). 16 April 2012. . Retrieved 2012-04-21.
- [65] "クリプトン | VOCALOID2 - キャラクター・ポータル・シリーズ" (<http://www.crypton.co.jp/mp/pages/prod/vocaloid/>). Crypton.co.jp. . Retrieved 2012-04-21.
- [66] G., Adrian. "LA's Anime Expo hosting Hatsune Miku's first US live performance on July 2nd" (<http://www.kawaiikkoiisugoi.com/2011/06/16/las-anime-expo-hosting-hatsune-mikus-first-us-live-performance-on-july-2nd/>). . Retrieved 20 April 2012.
- [67] ""We can invite Hatsune Miku in my room!", Part 2 (video)" (<http://www.youtube.com/watch?v=ZCYJu7KSqQM>). Youtube.com. 2011-09-07. . Retrieved 2012-04-21.
- [68] Firth, Niall (12 November 2010). "Japanese 3D singing hologram Hatsune Miku becomes nation's strangest pop star" (<http://www.dailymail.co.uk/sciencetech/article-1329040/Japanese-3D-singing-hologram-Hatsune-Miku-nations-biggest-pop-star.html>). London: Daily mail online. . Retrieved 29 April 2011.
- [69] "Techically incorrect: Tomorrow's Miley Cyrus? A hologram live in concert!" ([http://news.cnet.com/8301-17852\\_3-20022743-71.html](http://news.cnet.com/8301-17852_3-20022743-71.html)). . Retrieved 29 April 2011.
- [70] "Hatsune Miku – World is Mine Live in HD" (<http://www.youtube.com/watch?v=DTXO7KGHtjI&feature=related>). . Retrieved 29 April 2011.
- [71] "Burberry Beijing - Full Show" (<http://www.youtube.com/watch?v=9t5dCIuz2wY>). Youtube.com. . Retrieved 2012-04-21.
- [72] "Burberry lands in China" (<http://www.vogue.it/en/shows/fashion-events/2011/04/burberry-in-china>). . Retrieved June 14, 2011.

## Reference sources

- Hariharan P, 1996, *Optical Holography*, Cambridge University Press, ISBN 0-521-43965-5
- Hariharan P, 2002, *Basics of Holography*, Cambridge University Press, ISBN 0-521-00200-1
- Lipson A., Lipson SG, Lipson H, *Optical Physics*, 2011, Cambridge University Press, ISBN 978-0-521-49345-1

## Further reading

- *Lasers and holography: an introduction to coherent optics* W. E. Kock, Dover Publications (1981), ISBN 978-0-486-24041-1
- *Principles of holography* H. M. Smith, Wiley (1976), ISBN 978-0-471-80341-6
- G. Berger et al., *Digital Data Storage in a phase-encoded holographic memory system: data quality and security*, Proceedings of SPIE, Vol. 4988, p. 104–111 (2003)
- *Holographic Visions: A History of New Science* Sean F. Johnston, Oxford University Press (2006), ISBN 0-19-857122-4
- Saxby, Graham (2003). *Practical Holography, Third Edition*. Taylor and Francis. ISBN 978-0-7503-0912-7.
- *Three-Dimensional Imaging Techniques* Takanori Okoshi, Atara Press (2011), ISBN 978-0-9822251-4-1
- *Holographic Microscopy of Phase Microscopic Objects: Theory and Practice* Tatyana Tishko, Tishko Dmitry, Titar Vladimir, World Scientific (2010), ISBN 13 978-981-4289-54-2

## External links

- International Hologram Manufacturers Association (<http://www.ihma.org>)
- The nobel prize lecture of Denis Gabor (<http://nobelprize.org/physics/laureates/1971/gabor-autobio.html>)
- MIT's Spatial Imaging Group with papers about holographic theory and Holographic video (<http://www.media.mit.edu/spi/>)
- Medical Applications of Holograms ([http://www.integraf.com/a-holography\\_medical\\_applications.htm](http://www.integraf.com/a-holography_medical_applications.htm))
- How Stuff Works – holograms (<http://science.howstuffworks.com/hologram.htm>)
- Walker, John. (1992) "Holographic Art" (<http://www.artdesigncafe.com/holographic-art-1992>). *Glossary of Art, Architecture & Design since 1945*, 3rd. ed.
- Center for the Holographic Arts, New York – a non-profit organisation promoting holography (<http://holocenter.org>)
- Specular holography art site (<http://www.zintaglio.com>)
- Faster way to produce holographic tiles (<http://news.bbc.co.uk/2/hi/technology/7230258.stm>)
- Abrasion, hand-drawn holograms (<http://amasci.com/amateur/holo1.html>)
- Holoforum – A place to discuss holography (<http://holoforum.org/forum>)
- Animations demonstrating holography (<http://qed.wikina.org/holography/>) by QED
- U.S. Patent 3,506,327 (<http://www.google.com/patents?vid=3506327>) — "Wavefront reconstruction using a coherent reference beam" — E. N. Leith et al.

# Mathematics

**Mathematics** (from Greek μάθημα *máthēma*, "knowledge, study, learning") is the abstract study of topics encompassing quantity,<sup>[2]</sup> structure,<sup>[3]</sup> space,<sup>[2]</sup> change,<sup>[4][5]</sup> and other properties;<sup>[6]</sup> it has no generally accepted definition.<sup>[7][8]</sup>

Mathematicians seek out patterns<sup>[9][10]</sup> and formulate new conjectures. Mathematicians resolve the truth or falsity of conjectures by mathematical proof. The research required to solve mathematical problems can take years or even centuries of sustained inquiry. Since the pioneering work of Giuseppe Peano (1858–1932), David Hilbert (1862–1943), and others on axiomatic systems in the late 19th century, it has become customary to view mathematical research as establishing truth by rigorous deduction from appropriately chosen axioms and definitions. When those mathematical structures are good models of real phenomena, then mathematical reasoning can provide insight or predictions about nature.

Through the use of abstraction and logical reasoning, mathematics developed from counting, calculation, measurement, and the systematic study of the shapes and motions of physical objects. Practical mathematics has been a human activity for as far back as written records exist. Rigorous arguments first appeared in Greek mathematics, most notably in Euclid's *Elements*. Mathematics developed at a relatively slow pace until the Renaissance, when mathematical innovations interacting with new scientific discoveries led to a rapid increase in the rate of mathematical discovery that has continued to the present day.<sup>[11]</sup>

Galileo Galilei (1564–1642) said, "The universe cannot be read until we have learned the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible to comprehend a single word. Without these, one is wandering about in a dark labyrinth."<sup>[12]</sup> Carl Friedrich Gauss (1777–1855) referred to mathematics as "the Queen of the Sciences."<sup>[13]</sup> Benjamin Peirce (1809–1880) called mathematics "the science that draws necessary conclusions."<sup>[14]</sup> David Hilbert said of mathematics: "We are not speaking here of arbitrariness in any sense. Mathematics is not like a game whose tasks are determined by arbitrarily stipulated rules. Rather, it is a conceptual system possessing internal necessity that can only be so and by no means otherwise."<sup>[15]</sup> Albert Einstein (1879–1955) stated that "as far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."<sup>[16]</sup> French mathematician Claire Voisin states "There is creative drive in mathematics, it's all about movement trying to express itself."<sup>[17]</sup>

Mathematics is used throughout the world as an essential tool in many fields, including natural science, engineering, medicine, and the social sciences. Applied mathematics, the branch of mathematics concerned with application of mathematical knowledge to other fields, inspires and makes use of new mathematical discoveries, which has led to the development of entirely new mathematical disciplines, such as statistics and game theory. Mathematicians also engage in pure mathematics, or mathematics for its own sake, without having any application in mind. There is no clear line separating pure and applied mathematics, and practical applications for what began as pure mathematics are often discovered.<sup>[18]</sup>



Euclid, Greek mathematician, 3rd century BC, as imagined by Raphael in this detail from *The School of Athens*.<sup>[1]</sup>

## Etymology

The word *mathematics* comes from the Greek μάθημα (*máthēma*), which, in the ancient Greek language, means "what one learns", "what one gets to know", hence also "study" and "science", and in modern Greek just "lesson". The word *máthēma* is derived from μανθάνω (*manthano*), while the modern Greek equivalent is μαθαίνω (*mathaino*), both of which mean "to learn". In Greece, the word for "mathematics" came to have the narrower and more technical meaning "mathematical study", even in Classical times.<sup>[19]</sup> Its adjective is μαθηματικός (*mathēmatikós*), meaning "related to learning" or "studious", which likewise further came to mean "mathematical". In particular, μαθηματικὴ τέχνη (*mathēmatikḗ tékhnē*), Latin: *ars mathematica*, meant "the mathematical art".

In Latin, and in English until around 1700, the term *mathematics* more commonly meant "astrology" (or sometimes "astronomy") rather than "mathematics"; the meaning gradually changed to its present one from about 1500 to 1800. This has resulted in several mistranslations: a particularly notorious one is Saint Augustine's warning that Christians should beware of *mathematici* meaning astrologers, which is sometimes mistranslated as a condemnation of mathematicians.

The apparent plural form in English, like the French plural form *les mathématiques* (and the less commonly used singular derivative *la mathématique*), goes back to the Latin neuter plural *mathematica* (Cicero), based on the Greek plural τα μαθηματικά (*ta mathēmatiká*), used by Aristotle (384–322 BC), and meaning roughly "all things mathematical"; although it is plausible that English borrowed only the adjective *mathematic(al)* and formed the noun *mathematics* anew, after the pattern of physics and metaphysics, which were inherited from the Greek.<sup>[20]</sup> In English, the noun *mathematics* takes singular verb forms. It is often shortened to *maths* or, in English-speaking North America, *math*.<sup>[21]</sup>

## Definitions of mathematics

Aristotle defined mathematics as "the science of quantity", and this definition prevailed until the 18th century.<sup>[22]</sup> Starting in the 19th century, when the study of mathematics increased in rigor and began to address abstract topics such as group theory and projective geometry, which have no clear-cut relation to quantity and measurement, mathematicians and philosophers began to propose a variety of new definitions.<sup>[23]</sup> Some of these definitions emphasize the deductive character of much of mathematics, some emphasize its abstractness, some emphasize certain topics within mathematics. Today, no consensus on the definition of mathematics prevails, even among professionals.<sup>[7]</sup> There is not even consensus on whether mathematics is an art or a science.<sup>[8]</sup> A great many professional mathematicians take no interest in a definition of mathematics, or consider it undefinable.<sup>[7]</sup> Some just say, "Mathematics is what mathematicians do."<sup>[7]</sup>

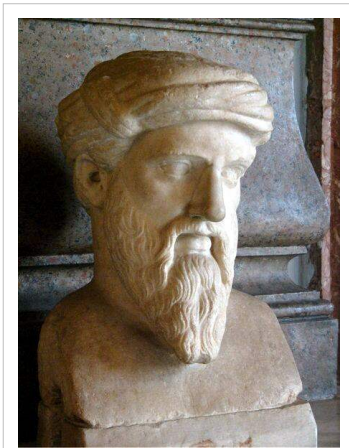
Three leading types of definition of mathematics are called *logicist*, *intuitionist*, and *formalist*, each reflecting a different philosophical school of thought.<sup>[24]</sup> All have severe problems, none has widespread acceptance, and no reconciliation seems possible.<sup>[24]</sup>

An early definition of mathematics in terms of logic was Benjamin Peirce's "the science that draws necessary conclusions" (1870).<sup>[25]</sup> In the *Principia Mathematica*, Bertrand Russell and Alfred North Whitehead advanced the philosophical program known as logicism, and attempted to prove that all mathematical concepts, statements, and principles can be defined and proven entirely in terms of symbolic logic. A logicist definition of mathematics is Russell's "All Mathematics is Symbolic Logic" (1903).<sup>[26]</sup>

Intuitionist definitions, developing from the philosophy of mathematician L.E.J. Brouwer, identify mathematics with certain mental phenomena. An example of an intuitionist definition is "Mathematics is the mental activity which consists in carrying out constructs one after the other."<sup>[24]</sup> A peculiarity of intuitionism is that it rejects some mathematical ideas considered valid according to other definitions. In particular, while other philosophies of mathematics allow objects that can be proven to exist even though they cannot be constructed, intuitionism allows only mathematical objects that one can actually construct.

Formalist definitions identify mathematics with its symbols and the rules for operating on them. Haskell Curry defined mathematics simply as "the science of formal systems".<sup>[27]</sup> A formal system is a set of symbols, or *tokens*, and some *rules* telling how the tokens may be combined into *formulas*. In formal systems, the word *axiom* has a special meaning, different from the ordinary meaning of "a self-evident truth". In formal systems, an axiom is a combination of tokens that is included in a given formal system without needing to be derived using the rules of the system.

## History



Greek mathematician Pythagoras (c. 570 – c. 495 BC), commonly credited with discovering the Pythagorean theorem.

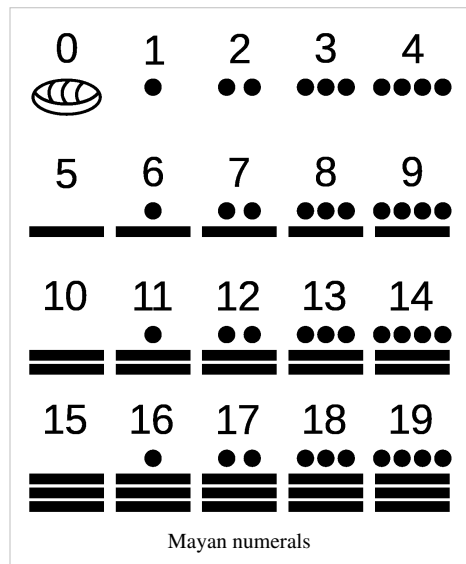
The evolution of mathematics might be seen as an ever-increasing series of abstractions, or alternatively an expansion of subject matter. The first abstraction, which is shared by many animals,<sup>[28]</sup> was probably that of numbers: the realization that a collection of two apples and a collection of two oranges (for example) have something in common, namely quantity of their members.

In addition to recognizing how to count *physical* objects, prehistoric peoples also recognized how to count *abstract* quantities, like time – days, seasons, years.<sup>[29]</sup> Elementary arithmetic (addition, subtraction, multiplication and division) naturally followed.

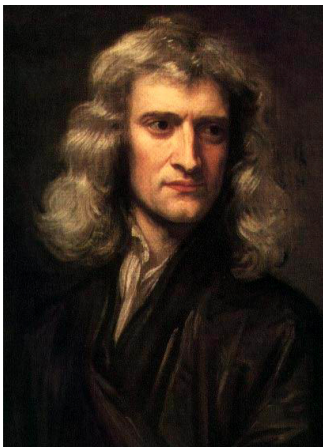
Since numeracy pre-dated writing, further steps were needed for recording numbers such as tallies or the knotted strings called quipu used by the Inca to store numerical data. Numeral systems have been many and diverse, with the first known written numerals created by Egyptians in Middle Kingdom texts such as the Rhind Mathematical Papyrus.

The earliest uses of mathematics were in trading, land measurement, painting and weaving patterns and the recording of time. More complex mathematics did not appear until around 3000 BC, when the Babylonians and Egyptians began using arithmetic, algebra and geometry for taxation and other financial calculations, for building and construction, and for astronomy.<sup>[30]</sup> The systematic study of mathematics in its own right began with the Ancient Greeks between 600 and 300 BC.<sup>[31]</sup>

Mathematics has since been greatly extended, and there has been a fruitful interaction between mathematics and science, to the benefit of both. Mathematical discoveries continue to be made today. According to Mikhail B. Sevryuk, in the January 2006 issue of the *Bulletin of the American Mathematical Society*, "The number of papers and books included in the *Mathematical Reviews* database since 1940 (the first year of operation of MR) is now more than 1.9 million, and more than 75 thousand items are added to the database each year. The overwhelming majority of works in this ocean contain new mathematical theorems and their proofs."<sup>[32]</sup>



## Inspiration, pure and applied mathematics, and aesthetics



Sir Isaac Newton (1643–1727), an inventor of infinitesimal calculus.

Mathematics arises from many different kinds of problems. At first these were found in commerce, land measurement, architecture and later astronomy; today, all sciences suggest problems studied by mathematicians, and many problems arise within mathematics itself. For example, the physicist Richard Feynman invented the path integral formulation of quantum mechanics using a combination of mathematical reasoning and physical insight, and today's string theory, a still-developing scientific theory which attempts to unify the four fundamental forces of nature, continues to inspire new mathematics.<sup>[33]</sup> Some mathematics is only relevant in the area that inspired it, and is applied to solve further problems in that area. But often mathematics inspired by one area proves useful in many areas, and joins the general stock of mathematical concepts. A distinction is often made between pure mathematics and applied mathematics. However pure mathematics topics often turn out to have applications, e.g. number theory in cryptography. This remarkable fact that even the "purest" mathematics often turns out to have practical applications is what Eugene

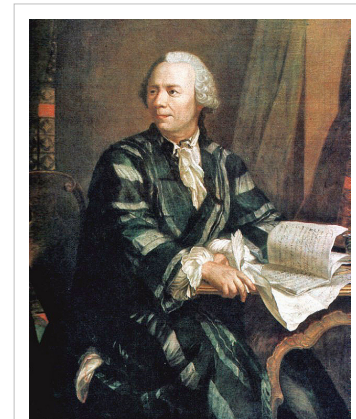
Wigner has called "the unreasonable effectiveness of mathematics".<sup>[34]</sup> As in most areas of study, the explosion of knowledge in the scientific age has led to specialization: there are now hundreds of specialized areas in mathematics and the latest Mathematics Subject Classification runs to 46 pages.<sup>[35]</sup> Several areas of applied mathematics have merged with related traditions outside of mathematics and become disciplines in their own right, including statistics, operations research, and computer science.

For those who are mathematically inclined, there is often a definite aesthetic aspect to much of mathematics. Many mathematicians talk about the *elegance* of mathematics, its intrinsic aesthetics and inner beauty. Simplicity and generality are valued. There is beauty in a simple and elegant proof, such as Euclid's proof that there are infinitely many prime numbers, and in an elegant numerical method that speeds calculation, such as the fast Fourier transform. G.H. Hardy in *A Mathematician's Apology* expressed the belief that these aesthetic considerations are, in themselves, sufficient to justify the study of pure mathematics. He identified criteria such as significance, unexpectedness, inevitability, and economy as factors that contribute to a mathematical aesthetic.<sup>[36]</sup> Mathematicians often strive to find proofs that are particularly elegant, proofs from "The Book" of God according to Paul Erdős.<sup>[37][38]</sup> The popularity of recreational mathematics is another sign of the pleasure many find in solving mathematical questions.



## Notation, language, and rigor

Most of the mathematical notation in use today was not invented until the 16th century.<sup>[39]</sup> Before that, mathematics was written out in words, a painstaking process that limited mathematical discovery.<sup>[40]</sup> Euler (1707–1783) was responsible for many of the notations in use today. Modern notation makes mathematics much easier for the professional, but beginners often find it daunting. It is extremely compressed: a few symbols contain a great deal of information. Like musical notation, modern mathematical notation has a strict syntax (which to a limited extent varies from author to author and from discipline to discipline) and encodes information that would be difficult to write in any other way.



Leonhard Euler, who created and popularized much of the mathematical notation used today

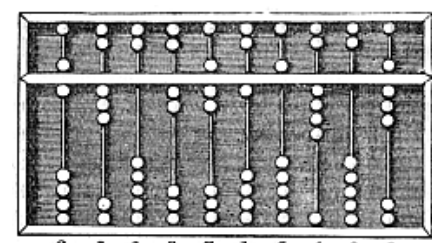
Mathematical language can be difficult to understand for beginners. Words such as *or* and *only* have more precise meanings than in everyday speech. Moreover, words such as *open* and *field* have been given specialized mathematical meanings. Technical terms such as *homeomorphism* and *integrable* have precise meanings in mathematics. Additionally, shorthand phrases such as *iff* for "if and only if" belong to mathematical jargon. There is a reason for special notation and technical vocabulary: mathematics requires more precision than everyday speech. Mathematicians refer to this precision of language and logic as "rigor".

Mathematical proof is fundamentally a matter of rigor. Mathematicians want their theorems to follow from axioms by means of systematic reasoning. This is to avoid mistaken "theorems", based on fallible intuitions, of which many instances have occurred in the history of the subject.<sup>[41]</sup> The level of rigor expected in mathematics has varied over time: the Greeks expected detailed arguments, but at the time of Isaac Newton the methods employed were less rigorous. Problems inherent in the definitions used by Newton would lead to a resurgence of careful analysis and formal proof in the 19th century. Misunderstanding the rigor is a cause for some of the common misconceptions of mathematics. Today, mathematicians continue to argue among themselves about computer-assisted proofs. Since large computations are hard to verify, such proofs may not be sufficiently rigorous.<sup>[42]</sup>

Axioms in traditional thought were "self-evident truths", but that conception is problematic. At a formal level, an axiom is just a string of symbols, which has an intrinsic meaning only in the context of all derivable formulas of an axiomatic system. It was the goal of Hilbert's program to put all of mathematics on a firm axiomatic basis, but according to Gödel's incompleteness theorem every (sufficiently powerful) axiomatic system has undecidable formulas; and so a final axiomatization of mathematics is impossible. Nonetheless mathematics is often imagined to be (as far as its formal content) nothing but set theory in some axiomatization, in the sense that every mathematical statement or proof could be cast into formulas within set theory.<sup>[43]</sup>

## Fields of mathematics

Mathematics can, broadly speaking, be subdivided into the study of quantity, structure, space, and change (i.e. arithmetic, algebra, geometry, and analysis). In addition to these main concerns, there are also subdivisions dedicated to exploring links from the heart of mathematics to other fields: to logic, to set theory (foundations), to the empirical mathematics of the various sciences (applied mathematics), and more recently to the rigorous study of uncertainty.



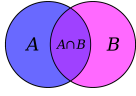
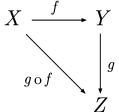
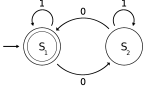
An abacus, a simple calculating tool used since ancient times.

## Foundations and philosophy

In order to clarify the foundations of mathematics, the fields of mathematical logic and set theory were developed. Mathematical logic includes the mathematical study of logic and the applications of formal logic to other areas of mathematics; set theory is the branch of mathematics that studies sets or collections of objects. Category theory, which deals in an abstract way with mathematical structures and relationships between them, is still in development. The phrase "crisis of foundations" describes the search for a rigorous foundation for mathematics that took place from approximately 1900 to 1930.<sup>[44]</sup> Some disagreement about the foundations of mathematics continues to the present day. The crisis of foundations was stimulated by a number of controversies at the time, including the controversy over Cantor's set theory and the Brouwer–Hilbert controversy.

Mathematical logic is concerned with setting mathematics within a rigorous axiomatic framework, and studying the implications of such a framework. As such, it is home to Gödel's incompleteness theorems which (informally) imply that any effective formal system that contains basic arithmetic, if *sound* (meaning that all theorems that can be proven are true), is necessarily *incomplete* (meaning that there are true theorems which cannot be proved *in that system*). Whatever finite collection of number-theoretical axioms is taken as a foundation, Gödel showed how to construct a formal statement that is a true number-theoretical fact, but which does not follow from those axioms. Therefore no formal system is a complete axiomatization of full number theory. Modern logic is divided into recursion theory, model theory, and proof theory, and is closely linked to theoretical computer science, as well as to category theory.

Theoretical computer science includes computability theory, computational complexity theory, and information theory. Computability theory examines the limitations of various theoretical models of the computer, including the most well-known model – the Turing machine. Complexity theory is the study of tractability by computer; some problems, although theoretically solvable by computer, are so expensive in terms of time or space that solving them is likely to remain practically unfeasible, even with the rapid advancement of computer hardware. A famous problem is the "**P = NP?**" problem, one of the Millennium Prize Problems.<sup>[45]</sup> Finally, information theory is concerned with the amount of data that can be stored on a given medium, and hence deals with concepts such as compression and entropy.

$p \Rightarrow q$			
Mathematical logic	Set theory	Category theory	Theory of computation

## Pure mathematics

### Quantity

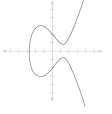

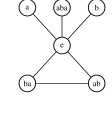
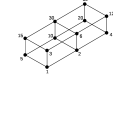
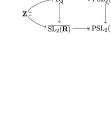
The study of quantity starts with numbers, first the familiar natural numbers and integers ("whole numbers") and arithmetical operations on them, which are characterized in arithmetic. The deeper properties of integers are studied in number theory, from which come such popular results as Fermat's Last Theorem. The twin prime conjecture and Goldbach's conjecture are two unsolved problems in number theory.

As the number system is further developed, the integers are recognized as a subset of the rational numbers ("fractions"). These, in turn, are contained within the real numbers, which are used to represent continuous quantities. Real numbers are generalized to complex numbers. These are the first steps of a hierarchy of numbers that goes on to include quaternions and octonions. Consideration of the natural numbers also leads to the transfinite numbers, which formalize the concept of "infinity". Another area of study is size, which leads to the cardinal numbers and then to another conception of infinity: the aleph numbers, which allow meaningful comparison of the size of infinitely large sets.

$1, 2, 3, \dots$	$\dots, -2, -1, 0, 1, 2 \dots$	$-2, \frac{2}{3}, 1.21$	$-e, \sqrt{2}, 3, \pi$	$2, i, -2 + 3i, 2e^{i\frac{4\pi}{3}}$
Natural numbers	Integers	Rational numbers	Real numbers	Complex numbers

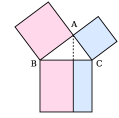
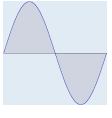
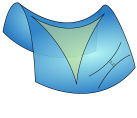
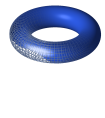
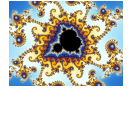
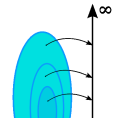
**Structure**

Many mathematical objects, such as sets of numbers and functions, exhibit internal structure as a consequence of operations or relations that are defined on the set. Mathematics then studies properties of those sets that can be expressed in terms of that structure; for instance number theory studies properties of the set of integers that can be expressed in terms of arithmetic operations. Moreover, it frequently happens that different such structured sets (or structures) exhibit similar properties, which makes it possible, by a further step of abstraction, to state axioms for a class of structures, and then study at once the whole class of structures satisfying these axioms. Thus one can study groups, rings, fields and other abstract systems; together such studies (for structures defined by algebraic operations) constitute the domain of abstract algebra. By its great generality, abstract algebra can often be applied to seemingly unrelated problems; for instance a number of ancient problems concerning compass and straightedge constructions were finally solved using Galois theory, which involves field theory and group theory. Another example of an algebraic theory is linear algebra, which is the general study of vector spaces, whose elements called vectors have both quantity and direction, and can be used to model (relations between) points in space. This is one example of the phenomenon that the originally unrelated areas of geometry and algebra have very strong interactions in modern mathematics. Combinatorics studies ways of enumerating the number of objects that fit a given structure.

$(1, 2, 3)$ $(1, 3, 2)$ $(2, 1, 3)$ $(2, 3, 1)$ $(3, 1, 2)$ $(3, 2, 1)$					
Combinatorics	Number theory	Group theory	Graph theory	Order theory	Algebra

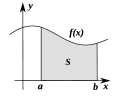
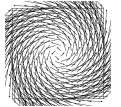
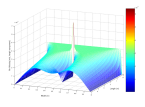
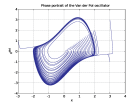

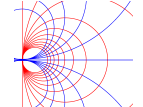
**Space**

The study of space originates with geometry – in particular, Euclidean geometry. Trigonometry is the branch of mathematics that deals with relationships between the sides and the angles of triangles and with the trigonometric functions; it combines space and numbers, and encompasses the well-known Pythagorean theorem. The modern study of space generalizes these ideas to include higher-dimensional geometry, non-Euclidean geometries (which play a central role in general relativity) and topology. Quantity and space both play a role in analytic geometry, differential geometry, and algebraic geometry. Convex and discrete geometry were developed to solve problems in number theory and functional analysis but now are pursued with an eye on applications in optimization and computer science. Within differential geometry are the concepts of fiber bundles and calculus on manifolds, in particular, vector and tensor calculus. Within algebraic geometry is the description of geometric objects as solution sets of polynomial equations, combining the concepts of quantity and space, and also the study of topological groups, which combine structure and space. Lie groups are used to study space, structure, and change. Topology in all its many ramifications may have been the greatest growth area in 20th century mathematics; it includes point-set topology, set-theoretic topology, algebraic topology and differential topology. In particular, instances of modern day topology are metrizable theory, axiomatic set theory, homotopy theory, and Morse theory. Topology also includes the now solved Poincaré conjecture, and the still unsolved areas of the Hodge conjecture. Other results in geometry and topology, including the four color theorem and Kepler conjecture, have been proved only with the help of computers.

					
Geometry	Trigonometry	Differential geometry	Topology	Fractal geometry	Measure theory

**Change**

Understanding and describing change is a common theme in the natural sciences, and calculus was developed as a powerful tool to investigate it. Functions arise here, as a central concept describing a changing quantity. The rigorous study of real numbers and functions of a real variable is known as real analysis, with complex analysis the equivalent field for the complex numbers. Functional analysis focuses attention on (typically infinite-dimensional) spaces of functions. One of many applications of functional analysis is quantum mechanics. Many problems lead naturally to relationships between a quantity and its rate of change, and these are studied as differential equations. Many phenomena in nature can be described by dynamical systems; chaos theory makes precise the ways in which many of these systems exhibit unpredictable yet still deterministic behavior.

					
Calculus	Vector calculus	Differential equations	Dynamical systems	Chaos theory	Complex analysis

**Applied mathematics**

Applied mathematics concerns itself with mathematical methods that are typically used in science, engineering, business, and industry. Thus, "applied mathematics" is a mathematical science with specialized knowledge. The term *applied mathematics* also describes the professional specialty in which mathematicians work on practical problems; as a profession focused on practical problems, *applied mathematics* focuses on the "formulation, study, and use of mathematical models" in science, engineering, and other areas of mathematical practice.

In the past, practical applications have motivated the development of mathematical theories, which then became the subject of study in pure mathematics, where mathematics is developed primarily for its own sake. Thus, the activity of applied mathematics is vitally connected with research in pure mathematics.

**Statistics and other decision sciences**

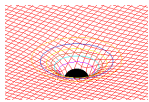
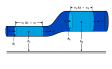
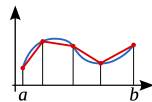
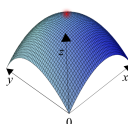

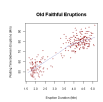




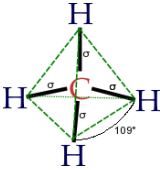

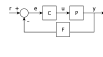
Applied mathematics has significant overlap with the discipline of statistics, whose theory is formulated mathematically, especially with probability theory. Statisticians (working as part of a research project) "create data that makes sense" with random sampling and with randomized experiments;<sup>[46]</sup> the design of a statistical sample or experiment specifies the analysis of the data (before the data be available). When reconsidering data from experiments and samples or when analyzing data from observational studies, statisticians "make sense of the data" using the art of modelling and the theory of inference – with model selection and estimation; the estimated models and consequential predictions should be tested on new data.<sup>[47]</sup>

Statistical theory studies decision problems such as minimizing the risk (expected loss) of a statistical action, such as using a procedure in, for example, parameter estimation, hypothesis testing, and selecting the best. In these traditional areas of mathematical statistics, a statistical-decision problem is formulated by minimizing an objective function, like expected loss or cost, under specific constraints: For example, designing a survey often involves minimizing the cost of estimating a population mean with a given level of confidence.<sup>[48]</sup> Because of its use of optimization, the mathematical theory of statistics shares concerns with other decision sciences, such as operations

research, control theory, and mathematical economics.<sup>[49]</sup>

**Computational mathematics**

Computational mathematics proposes and studies methods for solving mathematical problems that are typically too large for human numerical capacity. Numerical analysis studies methods for problems in analysis using functional analysis and approximation theory; numerical analysis includes the study of approximation and discretization broadly with special concern for rounding errors. Numerical analysis and, more broadly, scientific computing also study non-analytic topics of mathematical science, especially algorithmic matrix and graph theory. Other areas of computational mathematics include computer algebra and symbolic computation.

						
Mathematical physics	Fluid dynamics	Numerical analysis	Optimization	Probability theory	Statistics	Cryptography
						
Mathematical finance	Game theory	Mathematical biology	Mathematical chemistry	Mathematical economics	Control theory	

**Mathematics as profession**

Arguably the most prestigious award in mathematics is the Fields Medal,<sup>[50][51]</sup> established in 1936 and now awarded every four years. The Fields Medal is often considered a mathematical equivalent to the Nobel Prize.

The Wolf Prize in Mathematics, instituted in 1978, recognizes lifetime achievement, and another major international award, the Abel Prize, was introduced in 2003. The Chern Medal was introduced in 2010 to recognize lifetime achievement. These accolades are awarded in recognition of a particular body of work, which may be innovational, or provide a solution to an outstanding problem in an established field.

A famous list of 23 open problems, called "Hilbert's problems", was compiled in 1900 by German mathematician David Hilbert. This list achieved great celebrity among mathematicians, and at least nine of the problems have now been solved. A new list of seven important problems, titled the "Millennium Prize Problems", was published in 2000. A solution to each of these problems carries a \$1 million reward, and only one (the Riemann hypothesis) is duplicated in Hilbert's problems.



## Mathematics as science



Carl Friedrich Gauss, known as the "prince of mathematicians".<sup>[52]</sup>

Gauss referred to mathematics as "the Queen of the Sciences".<sup>[13]</sup> In the original Latin *Regina Scientiarum*, as well as in German *Königin der Wissenschaften*, the word corresponding to *science* means a "field of knowledge", and this was the original meaning of "science" in English, also. Of course, mathematics is in this sense a field of knowledge. The specialization restricting the meaning of "science" to *natural science* follows the rise of Baconian science, which contrasted "natural science" to scholasticism, the Aristotelean method of inquiring from first principles. Of course, the role of empirical experimentation and observation is negligible in mathematics, compared to natural sciences such as psychology, biology, or physics. Albert Einstein stated that "as far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."<sup>[16]</sup> More recently, Marcus du Sautoy has called mathematics "the Queen of Science ... the main driving force behind scientific discovery".<sup>[53]</sup>

Many philosophers believe that mathematics is not experimentally falsifiable, and thus not a science according to the definition of Karl Popper.<sup>[54]</sup> However, in the 1930s Gödel's incompleteness theorems convinced many mathematicians that mathematics cannot be reduced to logic alone, and Karl Popper concluded that "most mathematical theories are, like those of physics and biology, hypothetico-deductive: pure mathematics therefore turns out to be much closer to the natural sciences whose hypotheses are conjectures, than it seemed even recently."<sup>[55]</sup> Other thinkers, notably Imre Lakatos, have applied a version of falsificationism to mathematics itself.

An alternative view is that certain scientific fields (such as theoretical physics) are mathematics with axioms that are intended to correspond to reality. In fact, the theoretical physicist, J.M. Ziman, proposed that science is *public knowledge* and thus includes mathematics.<sup>[56]</sup> In any case, mathematics shares much in common with many fields in the physical sciences, notably the exploration of the logical consequences of assumptions. Intuition and experimentation also play a role in the formulation of conjectures in both mathematics and the (other) sciences. Experimental mathematics continues to grow in importance within mathematics, and computation and simulation are playing an increasing role in both the sciences and mathematics, weakening the objection that mathematics does not use the scientific method.

The opinions of mathematicians on this matter are varied. Many mathematicians feel that to call their area a science is to downplay the importance of its aesthetic side, and its history in the traditional seven liberal arts; others feel that to ignore its connection to the sciences is to turn a blind eye to the fact that the interface between mathematics and its applications in science and engineering has driven much development in mathematics. One way this difference of viewpoint plays out is in the philosophical debate as to whether mathematics is *created* (as in art) or *discovered* (as in science). It is common to see universities divided into sections that include a division of *Science and Mathematics*, indicating that the fields are seen as being allied but that they do not coincide. In practice, mathematicians are typically grouped with scientists at the gross level but separated at finer levels. This is one of many issues considered in the philosophy of mathematics.

## Notes

- [1] No likeness or description of Euclid's physical appearance made during his lifetime survived antiquity. Therefore, Euclid's depiction in works of art depends on the artist's imagination (*see Euclid*).
- [2] "mathematics, *n.*" (<http://oed.com/view/Entry/114974>). *Oxford English Dictionary*. Oxford University Press. 2012. . Retrieved June 16, 2012. "The science of space, number, quantity, and arrangement, whose methods involve logical reasoning and usually the use of symbolic notation, and which includes geometry, arithmetic, algebra, and analysis."
- [3] Kneebone, G.T. (1963). *Mathematical Logic and the Foundations of Mathematics: An Introductory Survey*. Dover. pp. 4 (<http://books.google.com/books?id=tCXxf4vbXCcC&pg=PA4>). ISBN 0486417123. "Mathematics...is simply the study of abstract structures, or formal patterns of connectedness."
- [4] LaTorre, Donald R., John W. Kenelly, Iris B. Reed, Laurel R. Carpenter, and Cynthia R Harris (2011). *Calculus Concepts: An Informal Approach to the Mathematics of Change*. Cengage Learning. pp. 2 (<http://books.google.com/books?id=1Ebu2Tij4QsC&pg=PA2>). ISBN 1439049572. "Calculus is the study of change—how things change, and how quickly they change."
- [5] Ramana (2007). *Applied Mathematics*. Tata McGraw–Hill Education. p. 2.10 (<http://books.google.com/books?id=XCRC6BeKhIIC&pg=SA2-PA10>). ISBN 0070667535. "The mathematical study of change, motion, growth or decay is calculus."
- [6] Ziegler, Günter M. (2011). "What Is Mathematics?". *An Invitation to Mathematics: From Competitions to Research*. Springer. pp. 7 (<http://books.google.com/books?id=9TATfteVeVYC&pg=PR7>). ISBN 3642195326.
- [7] Mura, Robert (Dec 1993). "Images of Mathematics Held by University Teachers of Mathematical Sciences". *Educational Studies in Mathematics* **25** (4): 375–385 (<http://www.jstor.org/stable/10.2307/3482762>).
- [8] Tobies, Renate and Helmut Neunzert (2012). *Iris Runge: A Life at the Crossroads of Mathematics, Science, and Industry*. Springer. pp. 9 (<http://books.google.com/books?id=EDm0eQqFUQ4C&pg=PA9>). ISBN 3034802293. "It is first necessary to ask what is meant by *mathematics* in general. Illustrious scholars have debated this matter until they were blue in the face, and yet no consensus has been reached about whether mathematics is a natural science, a branch of the humanities, or an art form."
- [9] Steen, L.A. (April 29, 1988). *The Science of Patterns* Science, 240: 611–616. And summarized at Association for Supervision and Curriculum Development (<http://www.ascd.org/publications/curriculum-handbook/409/chapters/The-Future-of-Mathematics-Education.aspx>), [www.ascd.org](http://www.ascd.org).
- [10] Devlin, Keith, *Mathematics: The Science of Patterns: The Search for Order in Life, Mind and the Universe* (Scientific American Paperback Library) 1996, ISBN 978-0-7167-5047-5
- [11] Eves
- [12] Marcus du Sautoy, *A Brief History of Mathematics: 1. Newton and Leibniz* (<http://www.bbc.co.uk/programmes/b00sr3fm>), BBC Radio 4, 27/09/2010.
- [13] Waltershausen
- [14] Peirce, p. 97.
- [15] Hilbert, D. (1919–20), *Natur und Mathematisches Erkennen: Vorlesungen, gehalten 1919–1920 in Göttingen. Nach der Ausarbeitung von Paul Bernays* (Edited and with an English introduction by David E. Rowe), Basel, Birkhäuser (1992).
- [16] Einstein, p. 28. The quote is Einstein's answer to the question: "how can it be that mathematics, being after all a product of human thought which is independent of experience, is so admirably appropriate to the objects of reality?" He, too, is concerned with *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*.
- [17] Claire Voisin, Artist of the Abstract (<http://www2.cnrs.fr/en/1402.htm>)
- [18] Peterson
- [19] Both senses can be found in Plato. Liddell and Scott, *s. voce μαθηματικός*
- [20] *The Oxford Dictionary of English Etymology, Oxford English Dictionary, sub "mathematics", "mathematic", "mathematics"*
- [21] "maths, *n.*" (<http://oed.com/view/Entry/114982>) and "math, *n.3*" (<http://oed.com/view/Entry/114962>). *Oxford English Dictionary*, on-line version (2012).
- [22] James Franklin, "Aristotelian Realism" in *Philosophy of Mathematics*, ed. A.D. Irvine, p. 104 (<http://books.google.com/books?id=mbn35b2ghgkC&pg=PA104#v=onepage&q&f=false>). Elsevier (2009).
- [23] Cajori, Florian (1893). *A History of Mathematics*. American Mathematical Society (1991 reprint). pp. 285–6 (<http://books.google.com/books?id=mGJRjC9fZgC&pg=PA285>). ISBN 0821821024.
- [24] Snapper, Ernst (September 1979), "The Three Crises in Mathematics: Logicism, Intuitionism, and Formalism", *Mathematics Magazine* **52** (4): 207–16, doi:10.2307/2689412, JSTOR 2689412.
- [25] Peirce, Benjamin (1882). *Linear Associative Algebra* (<http://books.google.com/books?id=De0GAAAAAYAAJ&pg=PA1#v=onepage&q&f=false>). p. 1. .
- [26] Bertrand Russell, *The Principles of Mathematics*, p. 5 ([http://books.google.com/books?id=kj0a\\_aV2mxIC&pg=PA5#v=onepage&q&f=false](http://books.google.com/books?id=kj0a_aV2mxIC&pg=PA5#v=onepage&q&f=false)). University Press, Cambridge (1903)
- [27] Curry, Haskell (1951). *Outlines of a Formalist Philosophy of Mathematics*. Elsevier. pp. 56 (<http://books.google.com/books?id=tZHrBQgp1bkC>). ISBN 0444533680.
- [28] S. Dehaene; G. Dehaene-Lambertz; L. Cohen (Aug 1998). "Abstract representations of numbers in the animal and human brain". *Trends in Neuroscience* **21** (8): 355–361. doi:10.1016/S0166-2236(98)01263-6. PMID 9720604.
- [29] See, for example, Raymond L. Wilder, *Evolution of Mathematical Concepts; an Elementary Study, passim*

- [30] Kline 1990, Chapter 1.
- [31] " *A History of Greek Mathematics: From Thales to Euclid* (<http://books.google.com/books?id=dmY3Vjix3kC&pg=PA1&dq&hl=en#v=onepage&q=&f=false>). Thomas Little Heath (1981). ISBN 0-486-24073-8
- [32] Sevryuk
- [33] Johnson, Gerald W.; Lapidus, Michel L. (2002). *The Feynman Integral and Feynman's Operational Calculus*. Oxford University Press. ISBN 0-8218-2413-9.
- [34] Wigner, Eugene (1960). "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" (<http://www.dartmouth.edu/~matc/MathDrama/reading/Wigner.html>). *Communications on Pure and Applied Mathematics* **13** (1): 1–14. doi:10.1002/cpa.3160130102. .
- [35] "Mathematics Subject Classification 2010" (<http://www.ams.org/mathscinet/msc/pdfs/classification2010.pdf>) (PDF). . Retrieved 2010-11-09.
- [36] Hardy, G.H. (1940). *A Mathematician's Apology*. Cambridge University Press. ISBN 0-521-42706-1.
- [37] Gold, Bonnie; Simons, Rogers A. (2008). *Proof and Other Dilemmas: Mathematics and Philosophy*. MAA.
- [38] Aigner, Martin; Ziegler, Günter M. (2001). *Proofs from The Book*. Springer. ISBN 3-540-40460-0.
- [39] Earliest Uses of Various Mathematical Symbols (<http://jeff560.tripod.com/mathsym.html>) (Contains many further references).
- [40] Kline, p. 140, on Diophantus; p. 261, on Vieta.
- [41] See *false proof* for simple examples of what can go wrong in a formal proof.
- [42] Ivars Peterson, *The Mathematical Tourist*, Freeman, 1988, ISBN 0-7167-1953-3. p. 4 "A few complain that the computer program can't be verified properly", (in reference to the Haken–Apple proof of the Four Color Theorem).
- [43] Patrick Suppes, *Axiomatic Set Theory*, Dover, 1972, ISBN 0-486-61630-4. p. 1, "Among the many branches of modern mathematics set theory occupies a unique place: with a few rare exceptions the entities which are studied and analyzed in mathematics may be regarded as certain particular sets or classes of objects."
- [44] Luke Howard Hodgkin & Luke Hodgkin, *A History of Mathematics*, Oxford University Press, 2005.
- [45] Clay Mathematics Institute ([http://www.claymath.org/millennium/P\\_vs\\_NP/](http://www.claymath.org/millennium/P_vs_NP/)), P=NP, claymath.org
- [46] Rao, C.R. (1997) *Statistics and Truth: Putting Chance to Work*, World Scientific. ISBN 981-02-3111-3
- [47] Like other mathematical sciences such as physics and computer science, statistics is an autonomous discipline rather than a branch of applied mathematics. Like research physicists and computer scientists, research statisticians are mathematical scientists. Many statisticians have a degree in mathematics, and some statisticians are also mathematicians.
- [48] Rao, C. R. (1981). "Foreword". In Arthanari, T. S.; Dodge, Yadolah. *Wiley Series in Probability and Mathematical Statistics*. Wiley. pp. vii–viii. ISBN 0-471-08073-X. MR607328.
- [49] Whittle (1994, pp. 10–11 and 14–18); Whittle, Peter (1994). "Almost home" ([http://www.statslab.cam.ac.uk/History/2history.html#6.\\_1966--72:\\_The\\_Churchill\\_Chair](http://www.statslab.cam.ac.uk/History/2history.html#6._1966--72:_The_Churchill_Chair)). In Kelly, F. P.. *Probability, statistics and optimisation: A Tribute to Peter Whittle* (previously "A realised path: The Cambridge Statistical Laboratory upto 1993 (revised 2002)" ed.). Chichester: John Wiley. pp. 1–28. ISBN 0-471-94829-2. .
- [50] "*The Fields Medal is now indisputably the best known and most influential award in mathematics.*" Monastyrsky
- [51] Riehm
- [52] Zeidler, Eberhard (2004). *Oxford User's Guide to Mathematics*. Oxford, UK: Oxford University Press. p. 1188. ISBN 0-19-850763-1.
- [53] Marcus du Sautoy, *A Brief History of Mathematics: 10. Nicolas Bourbaki* (<http://www.bbc.co.uk/programmes/b00stcgv>), BBC Radio 4, 01/10/2010.
- [54] Shasha, Dennis Elliot; Lazere, Cathy A. (1998). *Out of Their Minds: The Lives and Discoveries of 15 Great Computer Scientists*. Springer. p. 228.
- [55] Popper 1995, p. 56
- [56] Ziman

## References

- Courant, Richard and H. Robbins, *What Is Mathematics? : An Elementary Approach to Ideas and Methods*, Oxford University Press, USA; 2 edition (July 18, 1996). ISBN 0-19-510519-2.
- Einstein, Albert (1923). *Sidelights on Relativity: I. Ether and relativity. II. Geometry and experience (translated by G.B. Jeffery, D.Sc., and W. Perrett, Ph.D)*. (<http://searchworks.stanford.edu/view/1216826>). E.P. Dutton & Co., New York.
- du Sautoy, Marcus, *A Brief History of Mathematics* (<http://www.bbc.co.uk/podcasts/series/maths>), BBC Radio 4 (2010).
- Eves, Howard, *An Introduction to the History of Mathematics*, Sixth Edition, Saunders, 1990, ISBN 0-03-029558-0.
- Kline, Morris, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, USA; Paperback edition (March 1, 1990). ISBN 0-19-506135-7.



- Monastyrsky, Michael (2001) (PDF). *Some Trends in Modern Mathematics and the Fields Medal* ([http://www.fields.utoronto.ca/aboutus/FieldsMedal\\_Monastyrsky.pdf](http://www.fields.utoronto.ca/aboutus/FieldsMedal_Monastyrsky.pdf)). Canadian Mathematical Society. Retrieved 2006-07-28.
- Oxford English Dictionary, second edition, ed. John Simpson and Edmund Weiner, Clarendon Press, 1989, ISBN 0-19-861186-2.
- *The Oxford Dictionary of English Etymology*, 1983 reprint. ISBN 0-19-861112-9.
- Pappas, Theoni, *The Joy Of Mathematics*, Wide World Publishing; Revised edition (June 1989). ISBN 0-933174-65-9.
- Peirce, Benjamin (1881). Peirce, Charles Sanders. ed. "Linear associative algebra" (<http://books.google.com/?id=De0GAAAAYAAJ&pg=PA1&dq=Peirce+Benjamin+Linear+Associative+Algebra+&q=>). *American Journal of Mathematics* (Johns Hopkins University) **4** (1–4): 97–229. doi:10.2307/2369153. Corrected, expanded, and annotated revision with an 1875 paper by B. Peirce and annotations by his son, C. S. Peirce, of the 1872 lithograph ed. *Google Eprint* (<http://books.google.com/books?id=LQgPAAAAIAAJ&pg=PA221>) and as an extract, D. Van Nostrand, 1882, *Google Eprint* (<http://books.google.com/books?id=De0GAAAAYAAJ&printsec=frontcover>).
- Peterson, Ivars, *Mathematical Tourist, New and Updated Snapshots of Modern Mathematics*, Owl Books, 2001, ISBN 0-8050-7159-8.
- Popper, Karl R. (1995). "On knowledge". *In Search of a Better World: Lectures and Essays from Thirty Years*. Routledge. ISBN 0-415-13548-6.
- Riehm, Carl (August 2002). "The Early History of the Fields Medal" (<http://www.ams.org/notices/200207/comm-riehm.pdf>) (PDF). *Notices of the AMS* (AMS) **49** (7): 778–782.
- Sevryuk, Mikhail B. (January 2006). "Book Reviews" (<http://www.ams.org/bull/2006-43-01/S0273-0979-05-01069-4/S0273-0979-05-01069-4.pdf>) (PDF). *Bulletin of the American Mathematical Society* **43** (1): 101–109. doi:10.1090/S0273-0979-05-01069-4. Retrieved 2006-06-24.
- Waltershausen, Wolfgang Sartorius von (1856, repr. 1965). *Gauss zum Gedächtniss* (<http://www.amazon.de/Gauss-Gedächtnis-Wolfgang-Sartorius-Waltershausen/dp/3253017028>). Sändig Reprint Verlag H. R. Wohlwend. ISBN 3-253-01702-8. ISSN ASIN: B0000BN5SQ.

## Further reading

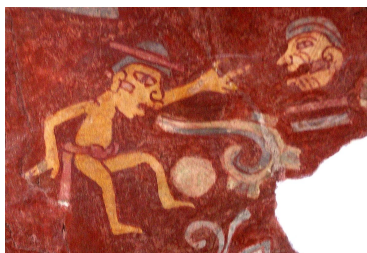
- Benson, Donald C., *The Moment of Proof: Mathematical Epiphanies*, Oxford University Press, USA; New Ed edition (December 14, 2000). ISBN 0-19-513919-4.
- Boyer, Carl B., *A History of Mathematics*, Wiley; 2nd edition, revised by Uta C. Merzbach, (March 6, 1991). ISBN 0-471-54397-7.—A concise history of mathematics from the Concept of Number to contemporary Mathematics.
- Davis, Philip J. and Hersh, Reuben, *The Mathematical Experience*. Mariner Books; Reprint edition (January 14, 1999). ISBN 0-395-92968-7.
- Gullberg, Jan, *Mathematics – From the Birth of Numbers*. W. W. Norton & Company; 1st edition (October 1997). ISBN 0-393-04002-X.
- Hazewinkel, Michiel (ed.), *Encyclopaedia of Mathematics*. Kluwer Academic Publishers 2000. — A translated and expanded version of a Soviet mathematics encyclopedia, in ten (expensive) volumes, the most complete and authoritative work available. Also in paperback and on CD-ROM, and online (<http://www.encyclopediaofmath.org>).
- Jourdain, Philip E. B., *The Nature of Mathematics*, in *The World of Mathematics*, James R. Newman, editor, Dover Publications, 2003, ISBN 0-486-43268-8.
- Maier, Annaliese, *At the Threshold of Exact Science: Selected Writings of Annaliese Maier on Late Medieval Natural Philosophy*, edited by Steven Sargent, Philadelphia: University of Pennsylvania Press, 1982.

## External links

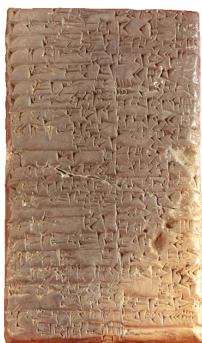
- Mathematics (<http://www.bbc.co.uk/programmes/p00545hk>) on *In Our Time* at the BBC. ( listen now ([http://www.bbc.co.uk/iplayer/console/p00545hk/In\\_Our\\_Time\\_Mathematics](http://www.bbc.co.uk/iplayer/console/p00545hk/In_Our_Time_Mathematics)))
- Free Mathematics books (<http://freebookcentre.net/SpecialCat/Free-Mathematics-Books-Download.html>) Free Mathematics books collection.
- Encyclopaedia of Mathematics online encyclopaedia from Springer (<http://www.encyclopediaofmath.org>), Graduate-level reference work with over 8,000 entries, illuminating nearly 50,000 notions in mathematics.
- HyperMath site at Georgia State University (<http://hyperphysics.phy-astr.gsu.edu/Hbase/hmat.html>)
- FreeScience Library (<http://www.freescience.info/mathematics.php>) The mathematics section of FreeScience library
- Rusin, Dave: *The Mathematical Atlas* (<http://www.math-atlas.org/>). A guided tour through the various branches of modern mathematics. (Can also be found at NIU.edu (<http://www.math.niu.edu/~rusin/known-math/index/index.html>).)
- Polyanin, Andrei: *EqWorld: The World of Mathematical Equations* (<http://eqworld.ipmnet.ru/>). An online resource focusing on algebraic, ordinary differential, partial differential (mathematical physics), integral, and other mathematical equations.
- Cain, George: Online Mathematics Textbooks (<http://www.math.gatech.edu/~cain/textbooks/onlinebooks.html>) available free online.
- Tricky (<http://www.tricky.org/>), Wiki-style site that is intended to develop into a large store of useful mathematical problem-solving techniques.
- Mathematical Structures (<http://math.chapman.edu/cgi-bin/structures?HomePage>), list information about classes of mathematical structures.
- Mathematician Biographies (<http://www-history.mcs.st-and.ac.uk/~history/>). The MacTutor History of Mathematics archive Extensive history and quotes from all famous mathematicians.
- *Metamath* (<http://metamath.org/>). A site and a language, that formalize mathematics from its foundations.
- Nrich (<http://www.nrich.maths.org/public/index.php>), a prize-winning site for students from age five from Cambridge University
- Open Problem Garden (<http://garden.irmacs.sfu.ca/>), a wiki of open problems in mathematics
- *Planet Math* (<http://planetmath.org/>). An online mathematics encyclopedia under construction, focusing on modern mathematics. Uses the Attribution-ShareAlike license, allowing article exchange with Wikipedia. Uses TeX markup.
- Some mathematics applets, at MIT (<http://www-math.mit.edu/daimp>)
- Weisstein, Eric et al.: *MathWorld: World of Mathematics* (<http://www.mathworld.com/>). An online encyclopedia of mathematics.
- Patrick Jones' Video Tutorials (<http://www.youtube.com/user/patrickJMT>) on Mathematics
- Citizendium: Theory (mathematics) ([http://en.citizendium.org/wiki/Theory\\_\(mathematics\)](http://en.citizendium.org/wiki/Theory_(mathematics))).
- du Sautoy, Marcus, *A Brief History of Mathematics* (<http://www.bbc.co.uk/podcasts/series/maths>), BBC Radio 4 (2010).
- MathOverflow (<http://mathoverflow.net/>) A Q&A site for research-level mathematics

# Language

---



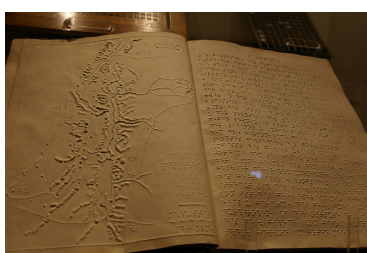
A mural in Teotihuacan, Mexico (ca. 200 AD) depicting a person emitting a speech scroll from his mouth, symbolizing speech.



Cuneiform is the first known form of written language, but spoken language predates writing by at least tens of thousands of years.



Two girls learning American Sign Language.



Braille writing represents language in a tactile form.

**Language** is the human capacity for acquiring and using complex systems of communication, and **a language** is any specific example of such a system. The scientific study of language is called linguistics. Any estimate of the precise number of languages in the world depends on a partly arbitrary distinction between languages and dialects. However, estimates vary between 6,000 and 7,000 languages in number. Natural languages are spoken or signed, but any language can be encoded into secondary media using auditory, visual, or tactile stimuli, for example, in graphic writing, braille, or whistling. This is because human language is modality-independent. When used as a general concept, "language" may refer to the cognitive ability to learn and use systems of complex communication, or to describe the set of rules that makes up these systems, or the set of utterances that can be produced from those rules. All languages rely on the process of semiosis to relate signs with particular meanings. Oral and sign languages

---

contain a phonological system that governs how symbols are used to form sequences known as words or morphemes, and a syntactic system that governs how words and morphemes are combined to form phrases and utterances.

Human language is unique because it has the properties of productivity, recursivity, and displacement, and because it relies entirely on social convention and learning. Its complex structure therefore affords a much wider range of possible expressions and uses than any known system of animal communication. Language is thought to have originated when early hominins started gradually changing their primate communication systems, acquiring the ability to form a theory of other minds and a shared intentionality.

This development is sometimes thought to have coincided with an increase in brain volume, and many linguists see the structures of language as having evolved to serve specific communicative and social functions. Language is processed in many different locations in the human brain, but especially in Broca's and Wernicke's areas. Humans acquire language through social interaction in early childhood, and children generally speak fluently when they are approximately three years old. The use of language is deeply entrenched in human culture. Therefore, in addition to its strictly communicative uses, language also has many social and cultural uses, such as signifying group identity, social stratification, as well as for social grooming and entertainment.

Languages evolve and diversify over time, and the history of their evolution can be reconstructed by comparing modern languages to determine which traits their ancestral languages must have had in order for the later stages to have occurred. A group of languages that descend from a common ancestor is known as a language family. The languages that are most spoken in the world today belong to the Indo-European family, which includes languages such as English, Spanish, Portuguese, Russian, and Hindi; the Sino-Tibetan languages, which include Mandarin Chinese, Cantonese, and many others; Semitic languages, which include Arabic, Amharic, and Hebrew; and the Bantu languages, which include Swahili, Zulu, Shona, and hundreds of other languages spoken throughout Africa. The general consensus is that between 50 and 90% of languages spoken today will probably have become extinct by the year 2100.<sup>[1][2]</sup>

## Definitions

The English word "language" derives ultimately from Indo-European *dn̥ǵʰwéh₂s* "tongue, speech, language" through Latin *lingua*, "language, tongue," and Old French *langage* "language."<sup>[3]</sup> The word is sometimes used to refer to codes, ciphers, and other kinds of artificially constructed communication systems such as those used for computer programming. A language in this sense is a system of signs for encoding and decoding information. This article specifically concerns the properties of natural human language as it is studied in the discipline of linguistics.

As an object of linguistic study, "language" has two primary meanings: an abstract concept, and a specific linguistic system, e.g. "French." The Swiss linguist Ferdinand de Saussure, who defined the modern discipline of linguistics, first explicitly formulated the distinction using the French word *langage* for language as a concept, *langue* as a specific instance of a language system, and *parole* for the concrete usage of speech in a particular language.<sup>[4]</sup>

When speaking of language as a general concept, definitions can be used which stress different aspects of the phenomenon.<sup>[5]</sup> These definitions also entail different approaches and understandings of language, and they inform different and often incompatible schools of linguistic theory.<sup>[6]</sup>

## Mental faculty, organ or instinct

One definition sees language primarily as the mental faculty that allows humans to undertake linguistic behaviour: to learn languages and to produce and understand utterances. This definition stresses the universality of language to all humans and it emphasizes the biological basis for the human capacity for language as a unique development of the human brain. Proponents of the view that the drive to language acquisition is innate in humans often argue that this is supported by the fact that all cognitively normal children raised in an environment where language is accessible will acquire language without formal instruction. Languages may even spontaneously develop in environments where

people live or grow up together without a common language, for example, creole languages and spontaneously developed sign languages such as Nicaraguan Sign Language. This view, which can be traced back to Kant and Descartes, often understands language to be largely innate, for example, in Chomsky's theory of Universal Grammar, or American philosopher Jerry Fodor's extreme innatist theory. These kinds of definitions are often applied by studies of language within a cognitive science framework and in neurolinguistics.<sup>[7][8]</sup>

### Formal symbolic system

Another definition sees language as a formal system of signs governed by grammatical rules of combination to communicate meaning. This definition stresses that human languages can be described as closed structural systems consisting of rules that relate particular signs to particular meanings.<sup>[9]</sup> This structuralist view of language was first introduced by Ferdinand de Saussure,<sup>[10]</sup> and his structuralism remains foundational for most approaches to language today.<sup>[11]</sup>

Some proponents of this view of language have advocated a formal approach which studies language structure by identifying its basic elements and then by formulating a formal account of the rules according to which the elements combine in order to form words and sentences. The main proponent of such a theory is Noam Chomsky, the originator of the generative theory of grammar, who has defined language as a particular set of sentences that can be generated from a particular set of rules.<sup>[12]</sup> Chomsky considers these rules to be an innate feature of the human mind and to constitute the essence of what language is.<sup>[13]</sup> Formal definitions of language are commonly used in formal logic, in formal theories of grammar, and in applied computational linguistics.<sup>[14][15]</sup>

### Tool for communication

Yet another definition sees language as a system of communication that enables humans to cooperate. This definition stresses the social functions of language and the fact that humans use it to express themselves and to manipulate objects in their environment. Functional theories of grammar explain grammatical structures by their communicative functions, and understand the grammatical structures of language to be the result of an adaptive process by which grammar was "tailored" to serve the communicative needs of its users.<sup>[16][17]</sup>

This view of language is associated with the study of language in pragmatic, cognitive, and interactive frameworks, as well as in socio-linguistics and linguistic anthropology. Functionalist theories tend to study grammar as dynamic phenomena, as structures that are always in the process of changing as they are employed by their speakers. This view places importance on the study of linguistic typology, or the classification of languages according to structural features, as it can be shown that processes of grammaticalization tend to follow trajectories that are partly dependent on typology. In the philosophy of language, these views are often associated with Wittgenstein's later works and with ordinary language philosophers such as Paul Grice, John Searle and J. L. Austin.<sup>[15]</sup>



Two men and a woman having a conversation in American Sign Language.

## What makes human language unique

Human language is unique in comparison to other forms of communication, such as those used by non-human animals. Communication systems used by other animals such as bees or non-human apes are closed systems that consist of a closed number of possible things that can be expressed.<sup>[18]</sup>

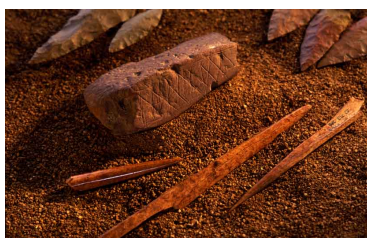
In contrast, human language is open-ended and productive, meaning that it allows humans to produce an infinite set of utterances from a finite set of elements and to create new words and sentences. This is possible because human language is based on a dual code, where a finite number of meaningless elements (e.g. sounds, letters or gestures) can be combined to form units of meaning (words and sentences).<sup>[19]</sup> Furthermore, the symbols and grammatical rules of any particular language are largely arbitrary, meaning that the system can only be acquired through social interaction.<sup>[20]</sup> The known systems of communication used by animals, on the other hand, can only express a finite number of utterances that are mostly genetically transmitted.<sup>[21]</sup>

Several species of animals have proven able to acquire forms of communication through social learning, such as the Bonobo Kanzi, which learned to express itself using a set of symbolic lexigrams. Similarly, many species of birds and whales learn their songs by imitating other members of their species. However, while some animals may acquire large numbers of words and symbols,<sup>[22]</sup> none have been able to learn as many different signs as is generally known by an average 4 year old human, nor have any acquired anything resembling the complex grammar of human language.<sup>[23]</sup>

Human languages also differ from animal communication systems in that they employ grammatical and semantic categories, such as noun and verb, present and past, to express exceedingly complex meanings.<sup>[23]</sup> Human language is also unique in having the property of recursivity: the way in which, for example, a noun phrase is able to contain another noun phrase (as in "[[the chimpanzee]'s lips]") or a clause is able to contain a clause (as in "[I see [the dog is running]]").<sup>[24]</sup> Human language is also the only known natural communication system that is *modality independent*, meaning that it can be used not only for communication through one channel or medium, but through several - for example, spoken language uses the auditive modality, whereas sign languages and writing use the visual modality, and braille writing uses the tactile modality.<sup>[25]</sup>

With regard to the meaning that it may convey and the cognitive operations that it builds on, human language is also unique in being able to refer to abstract concepts and to imagined or hypothetical events as well as events that took place in the past or may happen in the future. This ability to refer to events that are not at the same time or place as the speech event is called *displacement*, and while some animal communication systems can use displacement (such as the communication of bees that can communicate the location of sources of nectar that are out of sight), the degree to which it is used in human language is also considered unique.<sup>[19]</sup>

## Origin



75-80,000 year old artefacts from Blombos cave, South Africa, including a piece of ochre engraved with diagonal cross-hatch patterns, perhaps the oldest known example of symbols.





"The Tower of Babel" by Pieter Bruegel the Elder. Oil on board, 1563.

Humans have speculated about the origins of language throughout history. The Biblical myth of the Tower of Babel is one such account; other cultures have different stories of how language arose.<sup>[26]</sup>

Theories about the origin of language can be divided according to their basic assumptions. Some theories are based on the idea that language is so complex that one cannot imagine it simply appearing from nothing in its final form, but that it must have evolved from earlier pre-linguistic systems among our pre-human ancestors. These theories can be called continuity-based theories. The opposite viewpoint is that language is such a unique human trait that it cannot be compared to anything found among non-humans and that it must therefore have appeared suddenly in the transition from pre-hominids to early man. These theories can be defined as discontinuity-based. Similarly, theories based on Chomsky's Generative view of language see language mostly as an innate faculty that is largely genetically encoded, whereas functionalist theories see it as a system that is largely cultural, learned through social interaction.<sup>[27]</sup>

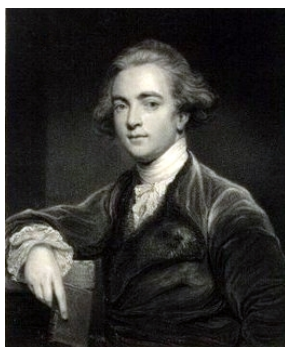
Currently, the only prominent proponent of a discontinuity-based theory of human language origins is linguist and philosopher Noam Chomsky. Chomsky proposes that "some random mutation took place, maybe after some strange cosmic ray shower, and it reorganized the brain, implanting a language organ in an otherwise primate brain."<sup>[28]</sup> While cautioning against taking this story too literally, Chomsky insists that "it may be closer to reality than many other fairy tales that are told about evolutionary processes, including language."<sup>[28]</sup>

Continuity-based theories are currently held by a majority of scholars, but they vary in how they envision this development. Those who see language as being mostly innate, for example, psychologist Steven Pinker, hold the precedents to be animal cognition,<sup>[8]</sup> whereas those who see language as a socially learned tool of communication, such as psychologist Michael Tomasello, see it as having developed from animal communication, either primate gestural or vocal communication to assist in cooperation.<sup>[21]</sup> Other continuity-based models see language as having developed from music, a view already espoused by Rousseau, Herder, Humboldt, and Charles Darwin. A prominent proponent of this view today is archaeologist Steven Mithen.<sup>[29]</sup>

Because the emergence of language is located in the early prehistory of man, the relevant developments have left no direct historical traces, and no comparable processes can be observed today. Theories that stress continuity often look at animals to see if, for example, primates display any traits that can be seen as analogous to what pre-human language must have been like. Alternatively, early human fossils can be inspected to look for traces of physical adaptation to language use or for traces of pre-linguistic forms of symbolic behaviour.<sup>[30]</sup>

It is mostly undisputed that pre-human australopithecines did not have communication systems significantly different from those found in great apes in general, but scholarly opinions vary as to the developments since the appearance of the genus *Homo* some 2.5 million years ago. Some scholars assume the development of primitive language-like systems (proto-language) as early as *Homo habilis* (2.3 million years ago), while others place the development of primitive symbolic communication only with *Homo erectus* (1.8 million years ago) or *Homo heidelbergensis* (0.6 million years ago), and the development of language proper with *Anatomically Modern Homo sapiens* with the Upper Paleolithic revolution less than 100,000 years ago.<sup>[31][32]</sup>

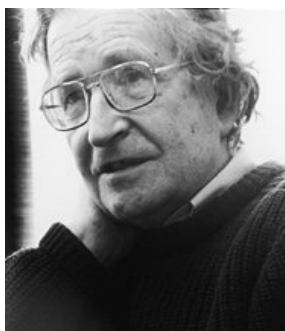
## The study of language



William Jones discovered the family relation between Latin and Sanskrit, laying the ground for the discipline of Historical linguistics.



Ferdinand de Saussure developed the structuralist approach to studying language.



Noam Chomsky is one of the most important linguistic theorists of the 20th century.

The study of language, linguistics, has been developing into a science since the first grammatical descriptions of particular languages in India more than 2000 years ago. Today, linguistics is a science that concerns itself with all aspects of language, examining it from all of the theoretical viewpoints described above.<sup>[33]</sup>

### Sub-disciplines

The academic study of language is conducted within many different disciplinary areas and from different theoretical angles, all of which inform modern approaches to linguistics. For example, descriptive linguistics examines the grammar of single languages, theoretical linguistics develops theories on how best to conceptualize and define the nature of language based on data from the various extant human languages, sociolinguistics studies how languages are used for social purposes informing in turn the study of the social functions of language and grammatical description, neurolinguistics studies how language is processed in the human brain and allows the experimental testing of theories, computational linguistics builds on theoretical and descriptive linguistics to construct computational models of language often aimed at processing natural language or at testing linguistic hypotheses, and historical linguistics relies on grammatical and lexical descriptions of languages to trace their individual histories and



reconstruct trees of language families by using the comparative method.<sup>[34]</sup>

## Early history

The formal study of language is often considered to have started in India with Pāṇini, the 5th century BC grammarian who formulated 3,959 rules of Sanskrit morphology. However, Sumerian scribes already studied the differences between Sumerian and Akkadian grammar around 1900 BC. Subsequent grammatical traditions developed in all of the ancient cultures that adopted writing.<sup>[35]</sup>

In the 17th century AD, the French Port-Royal Grammarians developed the idea that the grammars of all languages were a reflection of the universal basics of thought, and therefore that grammar was universal. In the 18th century, the first use of the comparative method by British philologist and expert on ancient India William Jones sparked the rise of comparative linguistics.<sup>[36]</sup> The scientific study of language was broadened from Indo-European to language in general by Wilhelm von Humboldt. Early in the 20th century, Ferdinand de Saussure introduced the idea of language as a static system of interconnected units, defined through the oppositions between them.<sup>[10]</sup>

By introducing a distinction between diachronic and synchronic analyses of language, he laid the foundation of the modern discipline of linguistics. Saussure also introduced several basic dimensions of linguistic analysis that are still fundamental in many contemporary linguistic theories, such as the distinctions between syntagm and paradigm, and the *Langue-parole* distinction, distinguishing language as an abstract system (*langue*), from language as a concrete manifestation of this system (*parole*).<sup>[37]</sup>

## Contemporary linguistics

In the 1960s, Noam Chomsky formulated the generative theory of language. According to this theory, the most basic form of language is a set of syntactic rules that is universal for all humans and which underlies the grammars of all human languages. This set of rules is called Universal Grammar; for Chomsky, describing it is the primary objective of the discipline of linguistics. Thus, he considered that the grammars of individual languages are only of importance to linguistics insofar as they allow us to deduce the universal underlying rules from which the observable linguistic variability is generated.<sup>[38]</sup>

In opposition to the formal theories of the generative school, functional theories of language propose that since language is fundamentally a tool, its structures are best analyzed and understood by reference to their functions. Formal theories of grammar seek to define the different elements of language and describe the way they relate to each other as systems of formal rules or operations, while functional theories seek to define the functions performed by language and then relate them to the linguistic elements that carry them out.<sup>[15][39]</sup> The framework of cognitive linguistics interprets language in terms of the concepts (which are sometimes universal, and sometimes specific to a particular language) which underlie its forms.<sup>[40]</sup> Cognitive linguistics is primarily concerned with how the mind creates meaning through language.

## Physiological and neural architecture of language and speech

Speaking is the default modality for language in all cultures. The production of spoken language depends on sophisticated capacities for controlling the lips, tongue and other components of the vocal apparatus, the ability to acoustically decode speech sounds, and the neurological apparatus required for acquiring and producing language.<sup>[41]</sup> The study of the genetic bases for human language is still on a fairly basic level, and the only gene that has been positively implied in language production is FOXP2, which may cause a kind of congenital language disorder if affected by mutations.<sup>[42]</sup>

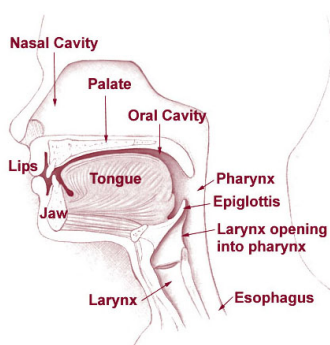
## The brain and language

The brain is the coordinating center of all linguistic activity; it controls both the production of linguistic cognition and of meaning and the mechanics of speech production. Nonetheless, our knowledge of the neurological bases for language is quite limited, though it has advanced considerably with the use of modern imaging techniques. The discipline of linguistics dedicated to studying the neurological aspects of language is called neurolinguistics.<sup>[43]</sup>

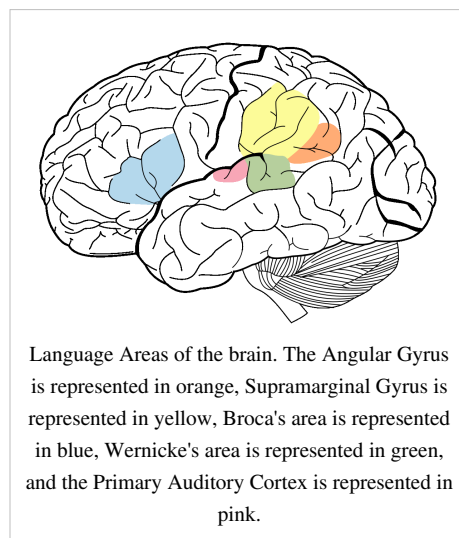
Early work in neurolinguistics involved the study of language in people with brain lesions, to see how lesions in specific areas affect language and speech. In this way, neuroscientists in the 19th century discovered that two areas in the brain are crucially implicated in language processing. The first area is Wernicke's area, which is located in the posterior section of the superior temporal gyrus in the dominant cerebral hemisphere. People with a lesion in this area of the brain develop receptive aphasia, a condition in which there is a major impairment of language comprehension, while speech retains a natural-sounding rhythm and a relatively normal sentence structure. The second area is Broca's area, located in the posterior inferior frontal gyrus of the dominant hemisphere. People with a lesion to this area develop expressive aphasia, meaning that they know "what they want to say, they just cannot get it out."<sup>[44]</sup> They are typically able to understand what is being said to them, but unable to speak fluently. Other symptoms that may be present in Broca's aphasia include problems with fluency, articulation, word-finding, word repetition, and producing and comprehending complex grammatical sentences, both orally and in writing. Those with this aphasia also exhibit ungrammatical speech and show inability to use syntactic information to determine the meaning of sentences. Both Broca's and Wernicke's aphasia also affect the use of sign language, in analogous ways to how they affect speech, with Broca's aphasia causing signers to sign slowly and with incorrect grammar, whereas a signer with Wernicke's aphasia will sign fluently, but make little sense to others and have difficulties comprehending others' signs. This shows that the impairment is specific to the ability to use language, not to the physiology used for speech production.<sup>[45][46]</sup>

With technological advances in the late 20th century, neurolinguists have also adopted non-invasive techniques such as functional magnetic resonance imaging (fMRI) and electrophysiology to study language processing in individuals without impairments.<sup>[43]</sup>

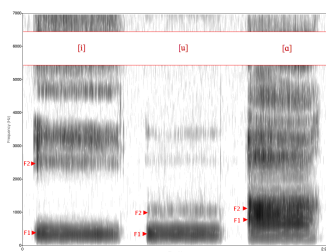
## Anatomy of speech



The human vocal tract.



Language Areas of the brain. The Angular Gyrus is represented in orange, Supramarginal Gyrus is represented in yellow, Broca's area is represented in blue, Wernicke's area is represented in green, and the Primary Auditory Cortex is represented in pink.



Spectrogram of American English vowels [i, u, ɑ] showing the formants  $f_1$  and  $f_2$



Real time MRI scan of a person speaking in Mandarin Chinese.

Spoken language relies on human physical ability to produce sound, which is a longitudinal wave propagated through the air at a frequency capable of vibrating the ear drum. This ability depends on the physiology of the human speech organs. These organs consist of the lungs, the voice box (larynx), and the upper vocal tract - the throat, the mouth, and the nose. By controlling the different parts of the speech apparatus, the airstream can be manipulated to produce different speech sounds.<sup>[47]</sup>

The sound of speech can be analyzed into a combination of segmental and suprasegmental elements. The segmental elements are those that follow each other in sequences, which are usually represented by distinct letters in alphabetic scripts, such as the Roman script. In free flowing speech, there are no clear boundaries between one segment and the next, nor usually are there any audible pauses between words. Segments therefore are distinguished by their distinct sounds which are a result of their different articulations, and they can be either vowels or consonants. Suprasegmental phenomena encompass such elements as stress, phonation type, voice timbre, and prosody or intonation, all of which may have effects across multiple segments.<sup>[48]</sup>

Consonants and vowel segments combine to form syllables, which in turn combine to form utterances; these can be distinguished phonetically as the space between two inhalations. Acoustically, these different segments are characterized by different formant structures, that are visible in a spectrogram of the recorded sound wave (See illustration of Spectrogram of the formant structures of three English vowels). Formants are the amplitude peaks in the frequency spectrum of a specific sound.<sup>[48][49]</sup>

Vowels are those sounds that have no audible friction caused by the narrowing or obstruction of some part of the upper vocal tract. They vary in quality according to the degree of lip aperture and the placement of the tongue within the oral cavity.<sup>[48]</sup> Vowels are called *close* when the lips are relatively closed, as in the pronunciation of the vowel [i] (English "ee"), or *open* when the lips are relatively open, as in the vowel [a] (English "ah"). If the tongue is located towards the back of the mouth, the quality changes, creating vowels such as [u] (English "oo"). The quality also changes depending on whether the lips are rounded as opposed to unrounded, creating distinctions such as that between [i] (unrounded front vowel such as English "ee") and [y] (rounded front vowel such as German "ü").<sup>[50]</sup>

Consonants are those sounds that have audible friction or closure at some point within the upper vocal tract. Consonant sounds vary by place of articulation, i.e. the place in the vocal tract where the airflow is obstructed, commonly at the lips, teeth, alveolar ridge, *palate*, *velum*, *uvula*, or *glottis*. Each place of articulation produces a different set of consonant sounds, which are further distinguished by manner of articulation, or the kind of friction,

whether full closure, in which case the consonant is called *occlusive* or *stop*, or different degrees of aperture creating *fricatives* and *approximants*. Consonants can also be either *voiced* or *unvoiced*, depending on whether the vocal cords are set in vibration by airflow during the production of the sound. Voicing is what separates English [s] in *bus* (unvoiced sibilant) from [z] in *buzz* (voiced sibilant).<sup>[51]</sup>

Some speech sounds, both vowels and consonants, involve release of air flow through the nasal cavity, and these are called *nasals* or *nasalized* sounds. Other sounds are defined by the way the tongue moves within the mouth: such as the l-sounds (called *laterals*, because the air flows along both sides of the tongue), and the r-sounds (called *rhotics*) that are characterized by how the tongue is positioned relative to the air stream.<sup>[49]</sup>

By using these speech organs, humans can produce hundreds of distinct sounds: some appear very often in the world's languages, whereas others are much more common in certain language families, language areas, or even specific to a single language.<sup>[52]</sup>

## Structure

When described as a system of symbolic communication, language is traditionally seen as consisting of three parts: signs, meanings, and a code connecting signs with their meanings. The study of the process of semiosis, how signs and meanings are combined, used, and interpreted is called *semiotics*. Signs can be composed of sounds, gestures, letters, or symbols, depending on whether the language is spoken, signed, or written, and they can be combined into complex signs, such as words and phrases. When used in communication, a sign is encoded and transmitted by a sender through a channel to a receiver who decodes it.<sup>[53]</sup>

Some of the properties that define human language as opposed to other communication systems are the arbitrariness of the linguistic sign, meaning that there is no predictable connection between a linguistic sign and its meaning, the duality of the linguistic system, meaning that linguistic structures are built by combining elements into larger structures that can be seen as layered, e.g. how sounds build words and words build phrases, the discreteness of the elements of language, meaning that the elements out of which linguistic signs are constructed are discrete units, e.g. sounds and words, that can be distinguished from each other and rearranged in different patterns, and the productivity of the linguistic system, meaning that the finite number of linguistic elements can be combined into a theoretically infinite number of combinations.<sup>[53]</sup>



Ancient Tamil inscription at Thanjavur

The rules by which signs can be combined to form words and phrases are called *syntax* or *grammar*. The meaning that is connected to individual signs, morphemes, words, phrases, and texts is called *semantics*.<sup>[54]</sup> The division of language into separate but connected systems of sign and meaning goes back to the first linguistic studies of de Saussure and is now used in almost all branches of linguistics.<sup>[55]</sup>

## Semantics

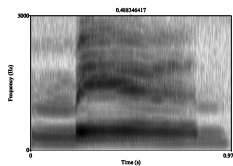
Languages express meaning by relating a sign form to a meaning, or its content. Sign forms must be something that can be perceived, for example, in sounds, images, or gestures, and then related to a specific meaning by social convention. Because the basic relation of meaning for most linguistic signs is based on social convention, linguistic signs can be considered arbitrary, in the sense that the convention is established socially and historically, rather than by means of a natural relation between a specific sign form and its meaning.

Thus, languages must have a vocabulary of signs related to specific meaning. The English sign "dog" denotes, for example, a member of the species *Canis familiaris*. In a language, the array of arbitrary signs connected to specific meanings is called the *lexicon*, and a single sign connected to a meaning is called a *lexeme*. Not all meanings in a

language are represented by single words. Often, semantic concepts are embedded in the morphology or syntax of the language in the form of grammatical categories.<sup>[56]</sup>

All languages contain the semantic structure of predication: a structure that predicates a property, state, or action. Traditionally, semantics has been understood to be the study of how speakers and interpreters assign truth values to statements, so that meaning is understood to be the process by which a predicate can be said to be true or false about an entity, e.g. "[x [is y]]" or "[x [does y]]." Recently, this model of semantics has been complemented with more dynamic models of meaning that incorporate shared knowledge about the context in which a sign is interpreted into the production of meaning. Such models of meaning are explored in the field of pragmatics.<sup>[56]</sup>

## Sounds and symbols



A spectrogram showing the sound of the spoken English word "man", which is written phonetically as [mæn]. Note that in flowing speech, there is no clear division between segments, only a smooth transition as the vocal apparatus moves.

The letter "wi" in the Hangul script.



The sign for "wi" in Korean Sign Language

Depending on modality, language structure can be based on systems of sounds (speech), gestures (sign languages), or graphic or tactile symbols (writing). The ways in which languages use sounds or signs to construct meaning are studied in phonology.<sup>[57]</sup> The study of how humans produce and perceive vocal sounds is called phonetics.<sup>[58]</sup> In spoken language, meaning is produced when sounds become part of a system in which some sounds can contribute to expressing meaning and others do not. In any given language, only a limited number of the many distinct sounds that can be created by the human vocal apparatus contribute to constructing meaning.<sup>[59]</sup>

Sounds as part of a linguistic system are called phonemes.<sup>[60]</sup> Phonemes are abstract units of sound, defined as the smallest units in a language that can serve to distinguish between the meaning of a pair of minimally different words, a so-called minimal pair. In English, for example, the words /bat/ [bat] and /pat/ [p<sup>h</sup>at] form a minimal pair, in which the distinction between /b/ and /p/ differentiates the two words, which have different meanings. However, each language contrasts sounds in different ways. For example, in a language that does not distinguish between voiced

and unvoiced consonants, the sounds [p] and [b] would be considered a single phoneme, and consequently, the two pronunciations would have the same meaning. Similarly, the English language does not distinguish phonemically between aspirated and non-aspirated pronunciations of consonants, as many other languages do: the unaspirated /p/ in /*spin*/ [spin] and the aspirated /p/ in /*pin*/ [p<sup>h</sup>in] are considered to be merely different ways of pronouncing the same phoneme (such variants of a single phoneme are called allophones), whereas in Mandarin Chinese, the same difference in pronunciation distinguishes between the words [p<sup>h</sup>á] "crouch" and [pá] "eight" (the accent above the á means that the vowel is pronounced with a high tone).<sup>[61]</sup>

All spoken languages have phonemes of at least two different categories, vowels and consonants, that can be combined to form syllables.<sup>[48]</sup> As well as segments such as consonants and vowels, some languages also use sound in other ways to convey meaning. Many languages, for example, use stress, pitch, duration, and tone to distinguish meaning. Because these phenomena operate outside of the level of single segments, they are called suprasegmental.<sup>[62]</sup> Some languages have only a few phonemes, for example, Rotokas and Pirahã language with 11 and 10 phonemes respectively, whereas languages like Taa may have as many as 141 phonemes.<sup>[61]</sup> In sign languages, the equivalent to phonemes (formerly called cheremes) are defined by the basic elements of gestures, such as hand shape, orientation, location, and motion, which correspond to manners of articulation in spoken language.<sup>[63]</sup>

Writing systems represent language using visual symbols, which may or may not correspond to the sounds of spoken language. The Latin alphabet (and those on which it is based or that have been derived from it) was originally based on the representation of single sounds, so that words were constructed from letters that generally denote a single consonant or vowel in the structure of the word. In syllabic scripts, such as the Inuktitut syllabary, each sign represents a whole syllable. In logographic scripts, each sign represents an entire word,<sup>[64]</sup> and will generally bear no relation to the sound of that word in spoken language.

Because all languages have a very large number of words, no purely logographic scripts are known to exist. Written language represents the way spoken sounds and words follow one after another by arranging symbols according to a pattern that follows a certain direction. The direction used in a writing system is entirely arbitrary and established by convention. Some writing systems use the horizontal axis (left to right as the Latin script or right to left as the Arabic script), while others such as traditional Chinese writing use the vertical dimension (from top to bottom). A few writing systems use opposite directions for alternating lines, and others, such as the ancient Maya script, can be written in either direction and rely on graphic cues to show the reader the direction of reading.<sup>[65]</sup>

In order to represent the sounds of the world's languages in writing, linguists have developed the International Phonetic Alphabet, designed to represent all of the discrete sounds that are known to contribute to meaning in human languages.<sup>[66]</sup>

## Grammar

Grammar is the study of how meaningful elements called *morphemes* within a language can be combined into utterances. Morphemes can either be *free* or *bound*. If they are free to be moved around within an utterance, they are usually called *words*, and if they are bound to other words or morphemes, they are called affixes. The way in which meaningful elements can be combined within a language is governed by rules. The rules for the internal structure of words are called morphology. The rules of the internal structure of phrases and sentences are called *syntax*.<sup>[67]</sup>

### Grammatical categories

Grammar can be described as a system of categories and a set of rules that determine how categories combine to form different aspects of meaning.<sup>[68]</sup> Languages differ widely in whether they are encoded through the use of categories or lexical units. However, several categories are so common as to be nearly universal. Such universal categories include the encoding of the grammatical relations of participants and predicates by grammatically distinguishing between their relations to a predicate, the encoding of temporal and spatial relations on predicates, and

a system of grammatical person governing reference to and distinction between speakers and addressees and those about whom they are speaking.<sup>[69]</sup>

### Word classes

Languages organize their parts of speech into classes according to their functions and positions relative to other parts. All languages, for instance, make a basic distinction between a group of words that prototypically denotes things and concepts and a group of words that prototypically denotes actions and events. The first group, which includes English words such as "dog" and "song," are usually called nouns. The second, which includes "run" and "sing," are called verbs. Another common category is the adjective: words that describe properties or qualities of nouns, such as "red" or "big." Word classes can be "open" if new words can continuously be added to the class, or relatively "closed" if there is a fixed number of words in a class. In English, the class of pronouns is closed, whereas the class of adjectives is open, since infinite numbers of adjectives can be constructed from verbs (e.g. "saddened") or nouns (e.g. with the -like suffix "noun-like"). In other languages such as Korean, the situation is the opposite, and new pronouns can be constructed, whereas the number of adjectives is fixed.<sup>[70]</sup>

The word classes also carry out differing functions in grammar. Prototypically, verbs are used to construct predicates, while nouns are used as arguments of predicates. In a sentence such as "Sally runs," the predicate is "runs," because it is the word that predicates a specific state about its argument "Sally." Some verbs such as "curse" can take two arguments, e.g. "Sally cursed John." A predicate that can only take a single argument is called *intransitive*, while a predicate that can take two arguments is called *transitive*.<sup>[71]</sup>

Many other word classes exist in different languages, such as conjunctions that serve to join two sentences, articles that introduce a noun, interjections such as "agh!" or "wow!", or ideophones that mimic the sound of some event. Some languages have positionals that describe the spatial position of an event or entity. Many languages have classifiers that identify countable nouns as belonging to a particular type or having a particular shape. For instance, in Japanese, the general noun classifier for humans is *nin* (人), and it is used for counting humans, whatever they are called:

*san-nin no gakusei* (三人の学生) lit. "3 human-classifier of student" — three students

For trees, it would be:

*san-bon no ki* (三本の木) lit. "3 classifier-for-long-objects of tree" — three trees

### Morphology

In linguistics, the study of the internal structure of complex words and the processes by which words are formed is called morphology. In most languages, it is possible to construct complex words that are built of several morphemes. For instance, the English word "unexpected" can be analyzed as being composed of the three morphemes "un-", "expect" and "-ed".<sup>[72]</sup>

Morphemes can be classified according to whether they are independent morphemes, so-called *roots*, or whether they can only co-occur attached to other morphemes. These bound morphemes or *affixes* can be classified according to their position in relation to the root: *prefixes* precede the root, *suffixes* follow the root, and *infixes* are inserted in the middle of a root. Affixes serve to modify or elaborate the meaning of the root. Some languages change the meaning of words by changing the phonological structure of a word, for example, the English word "run," which in the past tense is "ran." This process is called *ablaut*. Furthermore, morphology distinguishes between the process of *inflection*, which modifies or elaborates on a word, and the process of *derivation*, which creates a new word from an existing one. In English, the verb "sing" has the inflectional forms "singing" and "sung," which are both verbs, and the derivational form "singer," which is a noun derived from the verb with the agentive suffix "-er".<sup>[73][74]</sup>

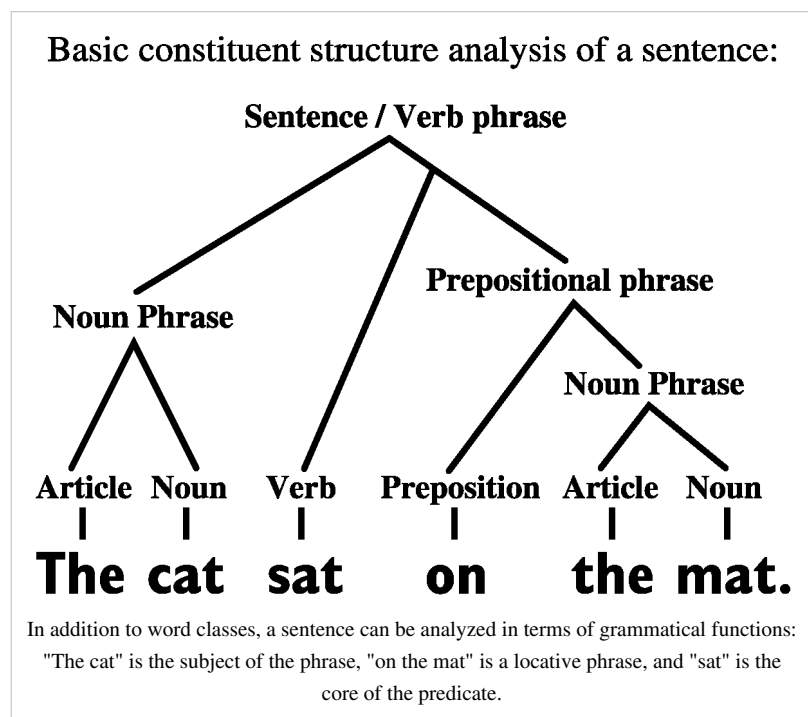
Languages differ widely in how much they rely on morphological processes of word formation. In some languages, for example, Chinese, there are no morphological processes, and all grammatical information is encoded syntactically by forming strings of single words. This type of morpho-syntax is often called isolating, or analytic,

because there is almost a full correspondence between a single word and a single aspect of meaning. Most languages have words consisting of several morphemes, but they vary in the degree to which morphemes are discrete units. In many languages, notably in most Indo-European languages, single morphemes may have several distinct meanings that cannot be analyzed into smaller segments. For example, in Latin, the word *bonus*, or "good," consists of the root *bon-*, meaning "good," and the suffix *-us*, which indicates masculine gender, singular number, and nominative case. These languages are called *fusional languages*, because several meanings may be fused into a single morpheme. The opposite type of fusional languages are agglutinative languages, which construct words by stringing morphemes together in chains, but with each morpheme as a discrete semantic unit. An example of such a language is Turkish, where for example, the word *evlerinizden*, or "from your houses," consists of the morphemes, *ev-ler-iniz-den* with the meanings *house-plural-your-from*. The languages that rely on morphology to the greatest extent are traditionally called polysynthetic languages. They may express the equivalent of an entire English sentence in a single word. For example, in the Yupik word *tuntussuqatarniksaitengqiggtuq*, which means "He had not yet said again that he was going to hunt reindeer," the word consists of the morphemes *tuntu-ssur-qatar-ni-ksaite-ngqiggte-uq* with the meanings, "reindeer-hunt-future-say-negation-again-third.person.singular.indicative," and except for the morpheme *tuntu*, or "reindeer," none of the other morphemes can appear in isolation.<sup>[75]</sup>

Many languages use morphology to cross-reference words within a sentence. This is sometimes called *agreement*. For example, in many Indo-European languages, adjectives must cross-reference the noun they modify in terms of number, case, and gender, so that the Latin adjective *bonus*, or "good," is inflected to agree with a noun that is masculine gender and singular. In many polysynthetic languages, verbs cross-reference their subjects and objects. In these types of languages, a single verb may include information that would require an entire sentence in English. For example, in the Basque phrase *ikusi nauzu*, or "you saw me," the past tense auxiliary verb *n-au-zu* (similar to English "do") agrees with both the subject (you) expressed by the *n-* prefix, and with the object (me) expressed by the *-zu* suffix. The sentence could be directly transliterated as "see you-did-me"<sup>[76]</sup>

## Syntax

Another way in which languages convey meaning is through the order of words within a sentence. The grammatical rules for how to produce new sentences from words that are already known is called syntax. The syntactical rules of a language determine why a sentence in English such as "I love you" is meaningful, but "\*love you I" is not.<sup>[77]</sup> Syntactical rules determine how word order and sentence structure is constrained, and how those constraints contribute to meaning.<sup>[78]</sup> For example, in English, the two sentences "the slaves were cursing the master" and "the master was cursing the slaves" mean different things, because the role of the grammatical subject is encoded by the noun being in front of the verb, and the role of object is encoded by the noun appearing after the verb. Conversely, in



the sentence "The cat sat on the mat," the role of subject is encoded by the noun appearing in front of the verb, and the role of object is encoded by the noun appearing after the verb. Conversely, in



Latin, both *Dominus servos vituperabat* and *Servos vituperabat dominus* mean "the master was reprimanding the slaves," because *servos*, or "slaves," is in the accusative case, showing that they are the grammatical object of the sentence, and *dominus*, or "master," is in the nominative case, showing that he is the subject.<sup>[79]</sup>

Latin uses morphology to express the distinction between subject and object, whereas English uses word order. Another example of how syntactic rules contribute to meaning is the rule of inverse word order in questions, which exists in many languages. This rule explains why when in English, the phrase "John is talking to Lucy" is turned into a question, it becomes "Who is John talking to?", and not "John is talking to who?". The latter example may be used as a way of placing special emphasis on "who," thereby slightly altering the meaning of the question. Syntax also includes the rules for how complex sentences are structured by grouping words together in units, called phrases, that can occupy different places in a larger syntactic structure. Sentences can be described as consisting of phrases connected in a tree structure, connecting the phrases to each other at different levels.<sup>[80]</sup> To the right is a graphic representation of the syntactic analysis of the English sentence "the cat sat on the mat". The sentence is analyzed as being constituted by a noun phrase, a verb, and a prepositional phrase; the prepositional phrase is further divided into a preposition and a noun phrase, and the noun phrases consist of an article and a noun.<sup>[81]</sup>

The reason sentences can be seen as being composed of phrases is because each phrase would be moved around as a single element if syntactic operations were carried out. For example, "the cat" is one phrase, and "on the mat" is another, because they would be treated as single units if a decision was made to emphasize the location by moving forward the prepositional phrase: "[And] on the mat, the cat sat".<sup>[81]</sup> There are many different formalist and functionalist frameworks that propose theories for describing syntactic structures, based on different assumptions about what language is and how it should be described. Each of them would analyze a sentence such as this in a different manner.<sup>[15]</sup>

### Typology: universals and diversity

Languages can be classified in relation to their grammatical types. Languages that belong to different families nonetheless often have features in common, and these shared features tend to correlate.<sup>[82]</sup> For example, languages can be classified on the basis of their basic word order, the relative order of the verb, and its constituents in a normal indicative sentence. In English, the basic order is SVO: "The snake(S) bit(V) the man(O)," whereas for example, the corresponding sentence in the Australian language Gamilaraay would be *ḍuyugu ṇama ḍayn yi:y* (Snake Man Bit), SOV.<sup>[83]</sup> Word order type is relevant as a typological parameter, because basic word order type corresponds with other syntactic parameters, such as the relative order of nouns and adjectives, or of the use of prepositions or postpositions. Such correlations are called implicational universals. For example, most (but not all) languages that are of the SOV type have postpositions rather than prepositions, and have adjectives before nouns.<sup>[84]</sup>

Through the study of various types of word order, it has been discovered that not all languages group the relations between actors and actions into Subject, Object and Verb, as English does. This type is called the nominative-accusative type. Some languages called ergative, Gamilaraay among them, distinguish between Agents and Patients. In English transitive clauses, both the subject of intransitive sentences ("I run") and transitive sentences ("I love you") are treated in the same way, shown here by the nominative pronoun *I*. In ergative languages, the single participant in an intransitive sentence, such as "I run," is treated the same as the patient in a transitive sentence, giving the equivalent of "me run" and "you love me." Only in transitive sentences would the equivalent of the pronoun "I" be used.<sup>[83]</sup> In this way the semantic roles can map onto the grammatical relations in different ways, grouping an intransitive subject either with Agents (accusative type) or Patients (ergative type) or even making each of the three roles differently, which is called the tripartite type.<sup>[85]</sup>

The shared features of languages which belong to the same typological class type may have arisen completely independently. Their co-occurrence might be due to the universal laws governing the structure of natural languages, "language universals," or they might be the result of languages evolving convergent solutions to the recurring communicative problems that humans use language to solve.<sup>[16]</sup>

## Social contexts of use and transmission

While all humans have the ability to learn any language, they only do so if they grow up in an environment in which language exists and is used by others. Language is therefore dependent on communities of speakers in which children learn language from their elders and peers and themselves transmit language to their own children. Languages are used by those who speak them to communicate and to solve a plethora of social tasks. Many aspects of language use can be seen to be adapted specifically to these purposes.<sup>[16]</sup> Due to the way in which language is transmitted between generations and within communities, language perpetually changes, diversifying into new languages or converging due to language contact. The process is similar to the process of evolution, where the process of descent with modification leads to the formation of a phylogenetic tree.<sup>[86]</sup>

However, languages differ from biological organisms in that they readily incorporate elements from other languages through the process of diffusion, as speakers of different languages come into contact. Humans also frequently speak more than one language, acquiring their first language or languages as children, or learning new languages as they grow up. Because of the increased language contact in the globalizing world, many small languages are becoming endangered as their speakers shift to other languages that afford the possibility to participate in larger and more influential speech communities.<sup>[87]</sup>

## Usage and meaning

The semantic study of meaning assumes that meaning is located in a relation between signs and meanings that are firmly established through social convention. However, semantics does not study the way in which social conventions are made and affect language. Rather, when studying the way in which words and signs are used, it is often the case that words have different meanings, depending on the social context of use. An important example of this is the process called deixis, which describes the way in which certain words refer to entities through their relation between a specific point in time and space when the word is uttered. Such words are, for example, the word, "I" (which designates the person speaking), "now" (which designates the moment of speaking), and "here" (which designates the time of speaking). Signs also change their meanings over time, as the conventions governing their usage gradually change. The study of how the meaning of linguistic expressions changes depending on context is called pragmatics. Deixis is an important part of the way that we use language to point out entities in the world.<sup>[88]</sup> Pragmatics is concerned with the ways in which language use is patterned and how these patterns contribute to meaning. For example, in all languages, linguistic expressions can be used not just to transmit information, but to perform actions. Certain actions are made only through language, but nonetheless have tangible effects, e.g. the act of "naming," which creates a new name for some entity, or the act of "pronouncing someone man and wife," which creates a social contract of marriage. These types of acts are called speech acts, although they can of course also be carried out through writing or hand signing.<sup>[89]</sup>

The form of linguistic expression often does not correspond to the meaning that it actually has in a social context. For example, if at a dinner table a person asks, "Can you reach the salt?", that is, in fact, not a question about the length of the arms of the one being addressed, but a request to pass the salt across the table. This meaning is implied by the context in which it is spoken; these kinds of effects of meaning are called conversational implicatures. These social rules for which ways of using language are considered appropriate in certain situations and how utterances are to be understood in relation to their context vary between communities, and learning them is a large part of acquiring communicative competence in a language.<sup>[90]</sup>

## Language acquisition

All healthy, normally-developing human beings learn to use language. Children acquire the language or languages used around them: whichever languages they receive sufficient exposure to during childhood. The development is essentially the same for children acquiring sign or oral languages.<sup>[91]</sup> This learning process is referred to as first-language acquisition, since unlike many other kinds of learning, it requires no direct teaching or specialized study. In *The Descent of Man*, naturalist Charles Darwin called this process "an instinctive tendency to acquire an art."<sup>[8]</sup>

First language acquisition proceeds in a fairly regular sequence, though there is a wide degree of variation in the timing of particular stages among normally-developing infants. From birth, newborns respond more readily to human speech than to other sounds. Around one month of age, babies appear to be able to distinguish between different speech sounds. Around six months of age, a child will begin babbling, producing the speech sounds or handshapes of the languages used around them. Words appear around the age of 12 to 18 months; the average vocabulary of an eighteen-month old child is around 50 words. A child's first utterances are holophrases (literally "whole-sentences"), utterances that use just one word to communicate some idea. Several months after a child begins producing words, she or he will produce two-word utterances, and within a few more months will begin to produce telegraphic speech, or short sentences that are less grammatically complex than adult speech, but that do show regular syntactic structure. From roughly the age of three to five years, a child's ability to speak or sign is refined to the point that it resembles adult language.<sup>[92]</sup>

Acquisition of second and additional languages can come at any age, through exposure in daily life or courses. Children learning a second language are more likely to achieve native-like fluency than adults, but in general, it is very rare for someone speaking a second language to pass completely for a native speaker. An important difference between first language acquisition and additional language acquisition is that the process of additional language acquisition is influenced by languages that the learner already knows.



All normal children acquire language if they are exposed to it in their first years of life, even in cultures where adults rarely address infants and toddlers directly.

## Language and culture

Languages, understood as the particular set of speech norms of a particular community, are also a part of the larger culture of the community that speaks them. Languages do not differ only in pronunciation, vocabulary, or grammar, but also through having different "cultures of speaking." Humans use language as a way of signalling identity with one cultural group and difference from others. Even among speakers of one language, several different ways of using the language exist, and each is used to signal affiliation with particular subgroups within a larger culture. Linguists and anthropologists, particularly sociolinguists, ethnolinguists, and linguistic anthropologists have specialized in studying how ways of speaking vary between speech communities.<sup>[93]</sup>



Arnold Lakhovsky, *The Conversation* (circa 1935)

Linguists use the term "varieties" to refer to the different ways of speaking a language. This term includes geographically or socioculturally defined dialects as well as the jargons or styles of subcultures. Linguistic anthropologists and sociologists of language define communicative style as the ways that language is used and understood within a particular culture.<sup>[94]</sup>

Because norms for language use are shared by members of a specific group, communicative style also becomes a way of displaying and constructing group identity. Linguistic differences may become salient markers of divisions between social groups, for example, speaking a language with a particular accent may imply membership of an ethnic minority or social class, one's area of origin, or status as a second language speaker. These kinds of differences are not part of the linguistic system, but are an important part of how language users use language as a social tool for constructing groups.<sup>[95]</sup>

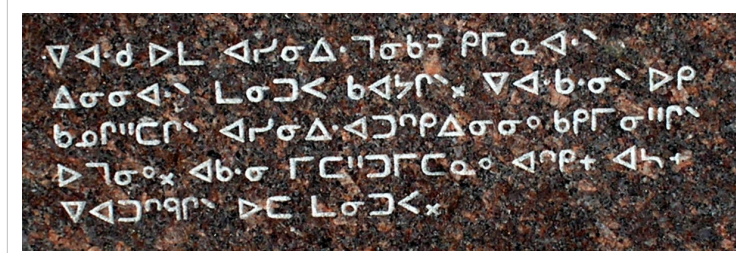
However, many languages also have grammatical conventions that signal the social position of the speaker in relation to others through the use of registers that are related to social hierarchies or divisions. In many languages, there are stylistic or even grammatical differences between the ways men and women speak, between age groups, or between social classes, just as some languages employ different words depending on who is listening. For example, in the Australian language Dyirbal, a married man must use a special set of words to refer to everyday items when speaking in the presence of his mother-in-law.<sup>[94]</sup> Some cultures, for example, have elaborate systems of "social deixis," or systems of signalling social distance through linguistic means.<sup>[96]</sup> In English, social deixis is shown mostly through distinguishing between addressing some people by first name and others by surname, and also in titles such as "Mrs.," "boy," "Doctor," or "Your Honor," but in other languages, such systems may be highly complex and codified in the entire grammar and vocabulary of the language. For instance, in several languages of east Asia, such as Thai, Burmese, and Javanese, different words are used according to whether a speaker is addressing someone of higher or lower rank than oneself in a ranking system with animals and children ranking the lowest and gods and members of royalty as the highest.<sup>[96]</sup>

## Writing, literacy and technology

Throughout history a number of different ways of representing language in graphic media have been invented. These are called writing systems.

The use of writing has made language even more useful to humans. It makes it possible to store large amounts of information outside of the human body and retrieve it again, and it allows communication across

distances that would otherwise be impossible. Many languages conventionally employ different genres, styles, and register in written and spoken language, and in some communities, writing traditionally takes place in an entirely different language than the one spoken. There is some evidence that the use of writing also has effects on the cognitive development of humans, perhaps because acquiring literacy generally requires explicit and formal education.<sup>[97]</sup>



An inscription of Swampy Cree using Canadian Aboriginal syllabics, an abugida developed by Christian missionaries for Indigenous Canadian languages

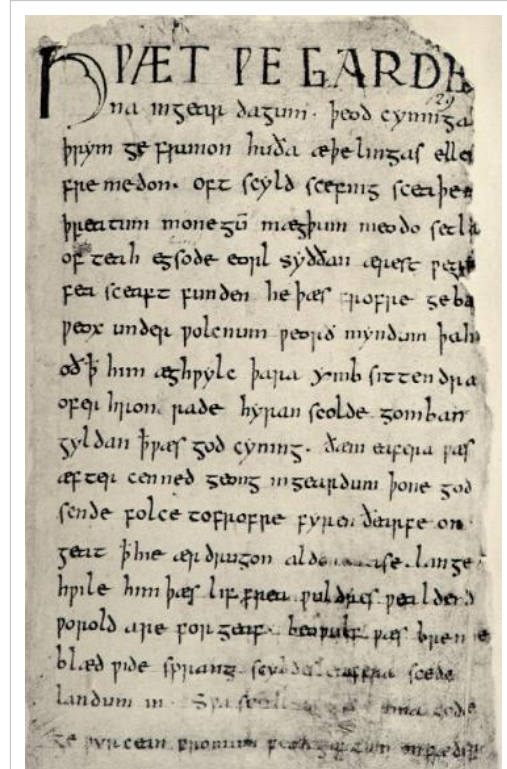
The invention of the first writing systems is roughly contemporary with the beginning of the Bronze Age in the late Neolithic period of the late 4th millennium BC. The Sumerian archaic cuneiform script and the Egyptian hieroglyphs are generally considered to be the earliest writing systems, both emerging out of their ancestral proto-literate symbol systems from 3400–3200 BC with the earliest coherent texts from about 2600 BC. It is generally agreed that Sumerian writing was an independent invention; however, it is debated whether Egyptian writing was developed completely independently of Sumerian, or was a case of cultural diffusion. A similar debate exists for the Chinese script, which developed around 1200 BC. The pre-Columbian Mesoamerican writing systems (including among others Olmec and Maya scripts) are generally believed to have had independent origins.<sup>[65]</sup>



## Language change

All languages change as speakers adopt or invent new ways of speaking and pass them on to other members of their speech community. Language change happens at all levels from the phonological level to the levels of vocabulary, morphology, syntax, and discourse. Even though language change is often initially evaluated negatively by speakers of the language who often consider changes to be "decay" or a sign of slipping norms of language usage, it is natural and inevitable.<sup>[98][99]</sup>

Changes may affect specific sounds or the entire phonological system. Sound change can consist of the replacement of one speech sound or phonetic feature by another, the complete loss of the affected sound, or even the introduction of a new sound in a place where there previously was none. Sound changes can be *conditioned* in which case a sound is changed only if it occurs in the vicinity of certain other sounds. Sound change is usually assumed to be *regular*, which means that it is expected to apply mechanically whenever its structural conditions are met, irrespective of any non-phonological factors. On the other hand, sound changes can sometimes be *sporadic*, affecting only one particular word or a few words, without any seeming regularity. Sometimes a simple change triggers a chain shift in which the entire phonological system is affected. This happened in the Germanic languages when the sound change known as Grimm's law affected all the stop consonants in the system. The original consonant \*b<sup>h</sup> became /b/ in the Germanic languages, the previous \*b in turn became /p/, and the previous \*p became /f/. The same process applied to all stop consonants and explains why Italic languages such as Latin have *pater* and *pisces*, whereas Germanic languages, like English, have *father* and *fish*.<sup>[100]</sup>



The first page of the Beowulf poem written in Old English in the early medieval period (800 - 1100 AD). Although old English language is the direct ancestor of modern English language, change has rendered it unintelligible to contemporary English speakers.

Another example is the Great Vowel Shift in English, which is the reason that the spelling of English vowels do not correspond well to their current pronunciation. This is because the vowel shift brought the already established orthography out of synchronization with pronunciation. Another source of sound change is the erosion of words as pronunciation gradually becomes increasingly indistinct and shortens words, leaving out syllables or sounds. This kind of change caused Latin *mea domina* to eventually become the French *madame* and American English *ma'am*.<sup>[101]</sup>

Change also happens in the grammar of languages as discourse patterns such as idioms or particular constructions become grammaticalized. This frequently happens when words or morphemes erode and the grammatical system is unconsciously rearranged to compensate for the lost element. For example, in some varieties of Caribbean Spanish the final /s/ has eroded away. Since Standard Spanish uses final /s/ in the morpheme marking the second person subject "you" in verbs, the Caribbean varieties now have to express the second person using the pronoun *tú*. This means that the sentence "what's your name" is *¿como te llamas?* ['komo te 'jamas] in Standard Spanish, but ['komo 'tu te 'jama] in Caribbean Spanish. The simple sound change has affected both morphology and syntax.<sup>[102]</sup> Another common cause of grammatical change is the gradual petrification of idioms into new grammatical forms, for example, the way the English "going to" construction lost its aspect of movement and in some varieties of English has almost become a full fledged future tense (e.g. *I'm gonna*).

Language change may be motivated by "language internal" factors, such as changes in pronunciation motivated by certain sounds being difficult to distinguish aurally or to produce, or because of certain patterns of change that cause certain rare types of constructions to drift towards more common types.<sup>[103]</sup> Other causes of language change are social, such as when certain pronunciations become emblematic of membership in certain groups, such as social classes, or with ideologies, and therefore are adopted by those who wish to identify with those groups or ideas. In this way, issues of identity and politics can have profound effects on language structure.<sup>[104]</sup>

## Language contact

One important source of language change is contact between different languages and resulting diffusion of linguistic traits between languages. Language contact occurs when speakers of two or more languages or varieties interact on a regular basis.<sup>[105]</sup> Multilingualism is likely to have been the norm throughout human history, and today, most people in the world are multilingual. Before the rise of the concept of the ethno-national state, monolingualism was characteristic mainly of populations inhabiting small islands. But with the ideology that made one people, one state, and one language the most desirable political arrangement, monolingualism started to spread throughout the world. Nonetheless, there are only 250 countries in the world corresponding to some 6000 languages, which means that most countries are multilingual and most languages therefore exist in close contact with other languages.<sup>[106]</sup>

When speakers of different languages interact closely, it is typical for their languages to influence each other. Through sustained language contact over long periods, linguistic traits diffuse between languages, and languages belonging to different families may converge to become more similar. In areas where many languages are in close contact, this may lead to the formation of language areas in which unrelated languages share a number of linguistic features. A number of such language areas have been documented, among them, the Balkan language area, the Mesoamerican language area, and the Ethiopian language area. Also, larger areas such as South Asia, Europe, and Southeast Asia have sometimes been considered language areas, because of widespread diffusion of specific areal features.<sup>[107][108]</sup>

Language contact may also lead to a variety of other linguistic phenomena, including language convergence, borrowing, and relexification (replacement of much of the native vocabulary with that of another language). In situations of extreme and sustained language contact, it may lead to the formation of new mixed languages that cannot be considered to belong to a single language family. One type of mixed language called pidgins occurs when adult speakers of two different languages interact on a regular basis, but in a situation where neither group learns to learn to speak the language of the other group fluently. In such a case, they will often construct a communication form that has traits of both languages, but which has a simplified grammatical and phonological structure. The language comes to contain mostly the grammatical and phonological categories that exist in both languages. Pidgin languages are defined by not having any native speakers, but only being spoken by people who have another language as their first language. But if a Pidgin language becomes the main language of a speech community, then eventually children will grow up learning the pidgin as their first language. As the generation of child learners grow up, the pidgin will often be seen to change its structure and acquire a greater degree of complexity. This type of language is generally called a creole language. An example of such mixed languages is Tok Pisin, the official language of Papua New-Guinea, which originally arose as a Pidgin based on English and Austronesian languages; others are Kreyòl ayisyen, the French based creole language spoken in Haiti, and Michif, a mixed language of Canada, based on the Native American language Cree and French.<sup>[109][110]</sup>

## Linguistic diversity

Language	Native speakers (in mil.) <sup>[111]</sup>
Mandarin	845
Spanish	329 <sup>[112]</sup>
English	328
Arabic languages	221
Hindi	182
Bengali	181
Portuguese	178
Russian	144
Japanese	122
German	90,3

A "living language" is simply one which is in wide use as a primary form of communication by a specific group of living people. The exact number of known living languages varies from 6,000 to 7,000, depending on the precision of one's definition of "language," and in particular, on how one defines the distinction between languages and dialects. As of 2009, SIL Ethnologue cataloged 6909 living human languages. The Ethnologue establishes linguistic groups based on studies of mutual intelligibility, and therefore often include more categories than more conservative classifications. For example, the Danish language that most scholars consider a single language with several dialects is classified as two distinct languages (Danish and Jutish) by the Ethnologue.<sup>[111]</sup>

The Ethnologue is also sometimes criticized for using cumulative data gathered over many decades, meaning that exact speaker numbers are frequently out of date, and some languages classified as living may have already become extinct. According to the Ethnologue, 389 (or nearly 6%) languages have more than a million speakers. These languages together account for 94% of the world's population, whereas 94% of the world's languages account for the remaining 6% of the global population. To the right is a table of the world's 10 most spoken languages with population estimates from the Ethnologue (2009 figures).<sup>[111]</sup>

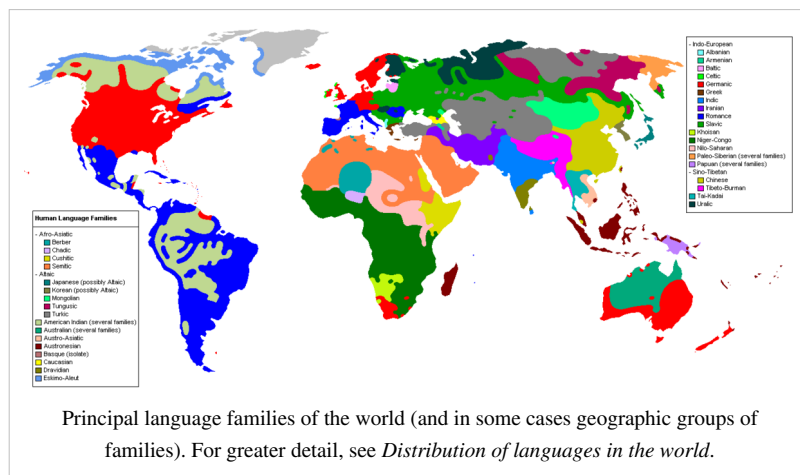
### Languages and dialects

There is no clear distinction between a language and a dialect, notwithstanding a famous aphorism attributed to linguist Max Weinreich that "a language is a dialect with an army and navy."<sup>[113]</sup> For example, national boundaries frequently override linguistic difference in determining whether two linguistic varieties are languages or dialects. Cantonese and Mandarin are, for example, often classified as "dialects" of Chinese, even though they are more different from each other than Swedish is from Norwegian. Before the Yugoslav civil war, Serbo-Croatian was considered a single language with two dialects, but now Croatian and Serbian are considered different languages and employ different writing systems. In other words, the distinction may hinge on political considerations as much as on cultural differences, distinctive writing systems, or degree of mutual intelligibility.<sup>[114]</sup>



## Language families of the World

The world's languages can be grouped into language families consisting of languages that can be shown to have common ancestry. Linguists currently recognize many hundreds of language families, although some of them can possibly be grouped into larger units as more evidence becomes available and in-depth studies are carried out. At present, there are also dozens of language isolates: languages that cannot be shown to be related to any other languages in the world. Among them is Basque, spoken in Europe, Zuni of New Mexico, P'urhépecha of Mexico, Ainu of Japan, Burushaski of Pakistan, and many others.



The language families of the world that have most speakers are the Indo-European languages, spoken by 46% of the world's population. This family includes major world languages like English, Spanish, Russian, and Hindustani (Hindi/Urdu). The Indo-European family achieved prevalence first during the Eurasian Migration Period (c. 400–800 AD), and subsequently through the European colonial expansion, which brought the Indo-European languages to a politically and often numerically dominant position in the Americas and much of Africa. The Sino-Tibetan languages are spoken by 21% of the world's population and include many of the languages of East Asia, including Mandarin Chinese, Cantonese, and hundreds of smaller languages.

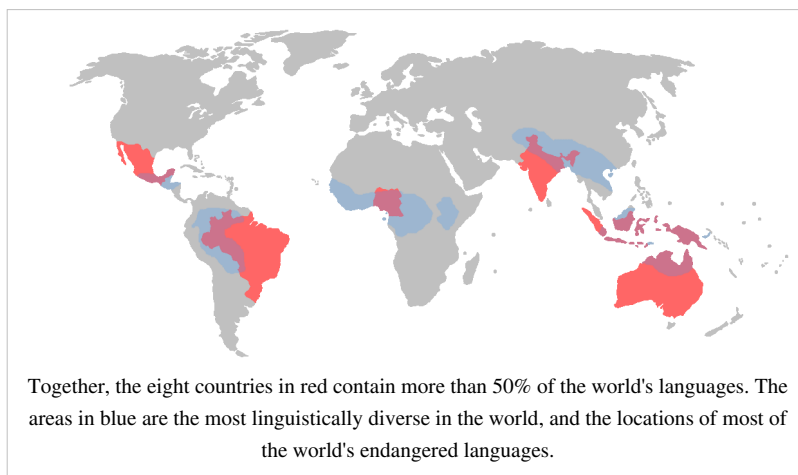
Africa is home to a large number of language families, the largest of which is the Niger-Congo language family, which includes such languages as Swahili, Shona, and Yoruba. Speakers of the Niger-Congo languages account for 6.4% of the world's population. A similar number of people speak the Afroasiatic languages, which include the populous Semitic languages such as Arabic, Hebrew language, and the languages of the Sahara region, such as the Berber languages and Hausa.

The Austronesian languages are spoken by 5.9% of the world's population and stretch from Madagascar to maritime Southeast Asia all the way to Oceania. It includes such languages as Malagasy, Māori, Samoan, and many of the indigenous languages of Indonesia and Taiwan. The Austronesian languages are considered to have originated in Taiwan around 3000 BC and spread through the Oceanic region through island-hopping, based on an advanced nautical technology. Other populous language families are the Dravidian languages of South Asia (among them Tamil and Telugu), the Turkic languages of Central Asia (such as Turkish), the Austroasiatic (Among them Khmer), and Tai–Kadai languages of Southeast Asia (including Thai).<sup>[115]</sup>

The areas of the world in which there is the greatest linguistic diversity, such as the Americas, Papua New Guinea, West Africa, and South-Asia, contain hundreds of small language families. These areas together account for the majority of the world's languages, though not the majority of speakers. In the Americas, some of the largest language families include the Quechumaran, Arawak, and Tupi-Guarani families of South America, the Uto-Aztecan, Oto-Manguean, and Mayan of Mesoamerica, and the Na-Dene and Algonquian language families of North America. In Australia, most indigenous languages belong to the Pama-Nyungan family, whereas Papua-New Guinea is home to a large number of small families and isolates, as well as a number of Austronesian languages.<sup>[115]</sup>

## Language endangerment

Language endangerment occurs when a language is at risk of falling out of use as its speakers die out or shift to speaking another language. Language loss occurs when the language has no more native speakers, and becomes a *dead language*. If eventually no one speaks the language at all, it becomes an *extinct language*. While languages have always gone extinct throughout human history, they are currently disappearing at an accelerated rate due to the processes of globalization and neo-colonialism, where the economically powerful languages dominate other languages.<sup>[1]</sup>



The more commonly spoken languages dominate the less commonly spoken languages and therefore, the less commonly spoken languages eventually disappear from populations. The total number of languages in the world is not known. Estimates vary depending on many factors. The general consensus is that there are between 6,000<sup>[2]</sup> and 7,000 languages currently spoken, and that between 50-90% of those will have become extinct by the year 2100.<sup>[1]</sup> The top 20 languages spoken by more than 50 million speakers each are spoken by 50% of the world's population, whereas many of the other languages are spoken by small communities, most of them with less than 10,000 speakers.<sup>[1]</sup>

The United Nations Educational, Scientific and Cultural Organization (UNESCO) operates with five levels of language endangerment: "safe," "vulnerable" (not spoken by children outside the home), "definitely endangered" (not spoken by children), "severely endangered" (only spoken by the oldest generations), and "critically endangered" (spoken by few members of the oldest generation, often semi-speakers). Notwithstanding claims that the world would be better off if most adopted a single common *lingua franca*, such as English or Esperanto, there is a general consensus that the loss of languages harms the cultural diversity of the world. It is a common belief, going back to the biblical narrative of the tower of Babel, that linguistic diversity causes political conflict,<sup>[26]</sup> but this belief is contradicted by the fact that many of the world's major episodes of violence have taken place in situations with low linguistic diversity, such as the Yugoslav and American Civil Wars, or the genocides of Nazi Germany and Rwanda, whereas many of the most stable political units have been highly multilingual.<sup>[116]</sup>

Many projects underway are aimed at preventing or slowing this loss by revitalizing endangered languages and promoting education and literacy in minority languages. Across the world, many countries have enacted specific legislation aimed at protecting and stabilizing the language of indigenous speech communities. A minority of linguists have argued that language loss is a natural process that should not be counteracted, and that documenting endangered languages for posterity is sufficient.<sup>[117]</sup>

## Notes

### Content notes

- [1] Austin & Sallabank (2011)
- [2] Moseley (2010)
- [3] "language". *The American Heritage Dictionary of the English Language* (3rd ed.). Boston: Houghton Mifflin Company. 1992.
- [4] Lyons (1981:2)
- [5] Lyons (1981:1–8)
- [6] Trask (2007:129–31)
- [7] Hauser & Fitch (2003)
- [8] Pinker (1994)
- [9] Trask 2007, p. 93
- [10] Saussure (1983)
- [11] Campbell (2001:96)
- [12] Trask 2007, p. 130
- [13] Chomsky (1957)
- [14] Trask (2007:93, 130)
- [15] Newmeyer (1998:3–6)
- [16] Evans & Levinson (2009)
- [17] Van Valin (2001)
- [18] Hockett (1960); Deacon (1997)
- [19] Trask (1999:1–5)
- [20] Trask (1999:9)
- [21] Tomasello (2008)
- [22] The Gorilla Koko reportedly uses as many as 1000 words in American Sign Language, and understands 2000 words of spoken English. There are some doubts about whether her use of signs is based in complex understanding or in simple conditioning (Candland (1993)).
- [23] Deacon (1997)
- [24] Hauser, Chomsky & Fitch (2002)
- [25] Trask (2007:165–66)
- [26] Haugen (1973)
- [27] Ulbaek (1998)
- [28] Chomsky 2000, p. 4
- [29] Fitch 2010, pp. 466–507
- [30] Fitch 2010, pp. 250–92
- [31] Foley 1997, pp. 70–74
- [32] Fitch 2010, pp. 292–3
- [33] Newmeyer (2005)
- [34] Trask (2007)
- [35] Campbell (2001:82–83)
- [36] Bloomfield 1914, p. 310
- [37] Clarke (1990:143–144)
- [38] Foley (1997:82–83)
- [39] "Functional grammar analyzes grammatical structure, as do formal and structural grammar; but it also analyzes the entire communicative situation: the purpose of the speech event, its participants, its discourse context. Functionalists maintain that the communicative situation motivates, constrains, explains, or otherwise determines grammatical structure, and that a structural or formal approaches not merely limited to an artificially restricted data base, but is inadequate even as a structural account. Functional grammar, then, differs from formal and structural grammar in that it purports not to model but to explain; and the explanation is grounded in the communicative situation." (Nichols (1984))
- [40] Croft & Cruse (2004:1)
- [41] Trask (1999:11–14; 105–113)
- [42] Fisher, Lai & Monaco (2003)
- [43] Lesser (1989:205–6)
- [44] Trask (1999:105–7)
- [45] Trask (1999:108)
- [46] Sandler & Lillo-Martin (2001:554)
- [47] MacMahon (1989:2)
- [48] MacMahon (1989:3)
- [49] International Phonetic Association (1999:3–8)
- [50] MacMahon (1989:11–15)
- [51] MacMahon (1989:6–11)

- [52] Ladefoged & Maddieson (1996)
- [53] Lyons (1981:17–24)
- [54] Trask (1999:35)
- [55] Lyons (1981:218–24)
- [56] Levinson (1983)
- [57] Goldsmith (1995)
- [58] International Phonetic Association (1999)
- [59] Ladefoged & Maddieson (1996)
- [60] International Phonetic Association (1999:27)
- [61] Trask (2007:214)
- [62] International Phonetic Association (1999:4)
- [63] Sandler & Lillo-Martin (2001:539–40)
- [64] Trask (2007:326)
- [65] Coulmas (2002)
- [66] Trask (2007:123)
- [67] Lyons (1981:103)
- [68] Allerton (1989)
- [69] Payne (1997)
- [70] Trask (2007:208)
- [71] Trask (2007:305)
- [72] Aronoff & Fudeman (2011:1–2)
- [73] Bauer (2003)
- [74] Haspelmath (2002)
- [75] Payne (1997:28–29)
- [76] Trask (2007:11)
- [77] The prefixed asterisk \* conventionally indicates that the sentence is ungrammatical, i.e. syntactically incorrect
- [78] Baker (2001:265)
- [79] Trask (2007:179)
- [80] Baker 2001, pp. 269–70
- [81] Trask (2007:218–19)
- [82] Nichols (1992); Comrie (1989)
- [83] Croft (2001:340)
- [84] Greenberg (1966)
- [85] Croft (2001:355)
- [86] Campbell (2004)
- [87] Austin & Sallabank (2011)
- [88] Levinson (1983:54–96)
- [89] Levinson (1983:226–78)
- [90] Levinson (1983:100–169)
- [91] Bonvillian, John D.; Michael D. Orlansky and Leslie Lazin Novack (December 1983). "Developmental milestones: Sign language acquisition and motor development". *Child Development* **54** (6): 1435–1445.
- [92] O'Grady, William; Cho, Sook Whan (2001). "First language acquisition". *Contemporary Linguistics: An Introduction* (fourth ed.). Boston: Bedford St. Martin's.
- [93] Duranti (2003)
- [94] Foley (1997)
- [95] Agha (2006)
- [96] Foley (1997:311–28)
- [97] Olson (1996)
- [98] Aitchison (2001)
- [99] Trask (1999:70)
- [100] Clackson (2007:27–33)
- [101] Aitchison (2001:112)
- [102] Zentella (2002:178)
- [103] Labov (1994)
- [104] Labov (2001)
- [105] Thomason (2001:1)
- [106] Romaine (2001:513)
- [107] Campbell (2002)
- [108] Aikhenvald (2001)

- [109] Thomason & Kaufman (1988); Thomason (2001)
- [110] Matras & Bakker (2003)
- [111] Lewis (2009)
- [112] Ethnologue's figure is based on numbers from before 1995. A more recent figure is 420 million ( "Primer estudio conjunto del Instituto Cervantes y el British Council sobre el peso internacional del español y del inglés" ([http://www.cervantes.es/sobre\\_instituto\\_cervantes/prensa/2012/noticias/nota-londres-palabra-por-palabra.htm](http://www.cervantes.es/sobre_instituto_cervantes/prensa/2012/noticias/nota-londres-palabra-por-palabra.htm)). Instituto Cervantes ([www.cervantes.es](http://www.cervantes.es)). )
- [113] Rickerson, E.M.. "What's the difference between dialect and language?" ([http://spinner.cofc.edu/linguist/archives/2005/08/whats\\_the\\_diffe.html?referrer=webcluster&](http://spinner.cofc.edu/linguist/archives/2005/08/whats_the_diffe.html?referrer=webcluster&)). *The Five Minute Linguist*. College of Charleston. . Retrieved 17 July 2011.
- [114] Lyons (1981:26)
- [115] Katzner (1999); Comrie (2009); Brown & Ogilvie (2008)
- [116] Austin & Sallabank (2011:10–11)
- [117] Ladefoged (1992)

## Citations

## References

- Agha, Agha (2006). *Language and Social Relations*. Cambridge University Press.
- Aikhenvald, Alexandra (2001). "Introduction". In Alexandra Y. Aikhenvald. *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press. pp. 1–26.
- Aitchison, Jean (2001). *Language Change: Progress or Decay?* (3rd (1st edition 1981) ed.). Cambridge, New York, Melbourne: Cambridge University Press.
- Allerton, D. J. (1989). "Language as Form and Pattern: Grammar and its Categories". In Collinge, N.E.. *An Encyclopedia of Language*. London:NewYork: Routledge.
- Aronoff, Mark; Fudeman, Kirsten (2011). *What is Morphology*. John Wiley & Sons.
- Austin, Peter K; Sallabank, Julia (2011). "Introduction". In Austin, Peter K; Sallabank, Julia. *Cambridge Handbook of Endangered Languages*. Cambridge University Press. ISBN 978-0-521-88215-6.
- Baker, Mark C. (2001). "Syntax". In Mark Aronoff. *The Handbook of Linguistics*. Blackwell. pp. 265–295.
- Bauer, Laurie (2003). *Introducing linguistic morphology* (2nd ed.). Washington, D.C.: Georgetown University Press. ISBN 0-87840-343-4.
- Bloomfield, Leonard (1914). *An introduction to the study of language*. New York: Henry Holt and Company.
- Brown, Keith; Ogilvie, Sarah, eds. (2008). *Concise Encyclopedia of Languages of the World*. Elsevier Science. ISBN 0080877745.
- Clackson, James (2007). *Indo-European Linguistics: An Introduction*. Cambridge University press.
- Campbell, Lyle (2002). "Areal linguistics". In Bernard Comrie, Neil J. Smelser and Paul B. Balte. *International Encyclopedia of Social and Behavioral Sciences*. Oxford: Pergamon. pp. 729–733.
- Campbell, Lyle (2004). *Historical Linguistics: an Introduction* (2nd ed.). Edinburgh and Cambridge, MA: Edinburgh University Press and MIT Press.
- Campbell, Lyle (2001). "The History of Linguistics". In Mark Aronoff. *The Handbook of Linguistics*. Blackwell. pp. 81–105.
- Candland, Douglas Keith (1993). *Feral Children and Clever Animals: Reflections on Human Nature* ([http://books.google.com/?id=RYG\\_CAWZZ4sC&pg=PA296&dq=koko+gorilla+operant+conditioning](http://books.google.com/?id=RYG_CAWZZ4sC&pg=PA296&dq=koko+gorilla+operant+conditioning)). Oxford University Press US. pp. 293–301. ISBN 0-19-510284-3.
- Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam (2000). *The Architecture of Language*. Oxford: Oxford University Press.
- Clarke, David S. (1990). *Sources of semiotic: readings with commentary from antiquity to the present*. Carbondale: Southern Illinois University Press.

- Comrie, Bernard (1989). *Language universals and linguistic typology: Syntax and morphology*. (2nd ed.). Oxford: Blackwell. ISBN 0-226-11433-3.
- Comrie, Bernard, ed. (2009). *The World's Major Languages*. New York: Routledge. ISBN 978-0-415-35339-7.
- Coulmas, Florian (2002). *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge University Press.
- Croft, William; Cruse, D. Alan (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Croft, William (2001). "Typology". In Mark Aronoff. *The Handbook of Linguistics*. Blackwell. pp. 81–105.
- Crystal, David (1997). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Deacon, Terrence (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: W.W. Norton & Company. ISBN 978-0-393-31754-1.
- Duranti, Alessandro (2003). "Language as Culture in U.S. Anthropology: Three Paradigms". *Current Anthropology* **44** (3): 323–348.
- Evans, Nicholas; Levinson, Stephen C. (2009). *The myth of language universals: Language diversity and its importance for cognitive science*. **32**. Behavioral and Brain Sciences. pp. 429–492.
- Fisher, Simon E.; Lai, Cecilia S.L.; Monaco, Anthony P. (2003). "Deciphering the Genetic Basis of Speech and Language Disorders". *Annual Review of Neuroscience* **26**: 57–80. doi:10.1146/annurev.neuro.26.041002.131144. PMID 12524432.
- Fitch, W. Tecumseh (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.
- Foley, William A. (1997). *Anthropological Linguistics: An Introduction*. Blackwell.
- Goldsmith, John A (1995). "Phonological Theory". In John A. Goldsmith. *The Handbook of Phonological Theory*. Blackwell Handbooks in Linguistics. Blackwell Publishers. ISBN 1-4051-5768-2.
- Greenberg, Joseph (1966). *Language Universals: With Special Reference to Feature Hierarchies*. The Hague: Mouton & Co.
- Haspelmath, Martin (2002). *Understanding morphology*. London: Arnold, Oxford University Press. (pbk)
- Haugen, Einar (1973). "The Curse of Babel". *Daedalus* **102** (3, Language as a Human Problem): 47–57.
- Hauser, Marc D.; Chomsky, Noam; Fitch, W. Tecumseh (2002). "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?". *Science* **22** **298** (5598): 1569–1579.
- Hauser, Marc D.; Fitch, W. Tecumseh (2003). "What are the uniquely human components of the language faculty?" (<http://www.isrl.uiuc.edu/~amag/langev/paper/hauser03whatAre.html>). In M.H. Christiansen and S. Kirby. *Language Evolution: The States of the Art*. Oxford University Press.
- Hockett, Charles F. (1960). "Logical considerations in the study of animal communication". In W.E. Lanyon. *Animals sounds and animal communication*. pp. 392–430.
- International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press. ISBN 0-521-65236-7 (hb); ISBN 0-521-63751-1 (pb).
- Katzner, K (1999). *The Languages of the World*. New York: Routledge.
- Labov, William (1994). *Principles of Linguistic Change vol.I Internal Factors*. Blackwell.
- Labov, William (2001). *Principles of Linguistic Change vol.II Social Factors*. Blackwell.
- Ladefoged, Peter (1992). "Another view of endangered languages". *Language* **68** (4): 809–811.
- Ladefoged, Ian; Maddieson (1996). *The sounds of the world's languages*. Oxford: Blackwell. pp. 329–330. ISBN 0-631-19815-6.

- Lesser, Ruth (1989). "Language in the Brain: Neurolinguistics". In Collinge, N.E.. *An Encyclopedia of Language*. London:NewYork: Routledge.
- Levinson, Stephen C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Lewis, M. Paul (ed.) (2009). "Ethnologue: Languages of the World, Sixteenth edition" ([http://www.ethnologue.com/ethno\\_docs/distribution.asp?by=size](http://www.ethnologue.com/ethno_docs/distribution.asp?by=size)). Dallas, Tex.: SIL International.
- Lyons, John (1981). *Language and Linguistics*. Cambridge University Press. ISBN 0-521-29775-3.
- MacMahon, M.K.C. (1989). "Language as available sound:Phonetics". In Collinge, N.E.. *An Encyclopedia of Language*. London:NewYork: Routledge.
- Matras, Yaron; Bakker, Peter, eds. (2003). *The Mixed Language Debate: Theoretical and Empirical Advances*. Berlin: Walter de Gruyter. ISBN 3-11-017776-5.
- Moseley, Christopher, ed. (2010). *Atlas of the World's Languages in Danger, 3rd edition*. (<http://www.unesco.org/culture/en/endangeredlanguages/atlas>). Paris: UNESCO Publishing.
- Newmeyer, Frederick J. (2005). *The History of Linguistics* (<http://www.lsadc.org/info/ling-fields-history.cfm>). Linguistic Society of America. ISBN 0-415-11553-1.
- Newmeyer, Frederick J. (1998). *Language Form and Language Function* ([http://www.isc.cnrs.fr/FN\\_chapter1.pdf](http://www.isc.cnrs.fr/FN_chapter1.pdf)). Cambridge,MA: MIT Press.
- Nichols, Johanna (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press. ISBN 0-226-58057-1.
- Nichols, Johanna (1984). "Functional Theories of Grammar". *Annual Review of Anthropology* **13**: 97–117.
- Olson, David R. (1996). "Language and Literacy: what writing does to Language and Mind". *Annual Review of Applied Linguistics* **16**: 3–13. doi:10.1017/S0267190500001392.
- Payne, Thomas Edward (1997). *Describing morphosyntax: a guide for field linguists* ([http://books.google.com/books?id=LC3DfjWfCiwC&pg=PA239&lpg=PA239&dq="perfect+aspect"+"perfective+aspect"&ct=result#PPA238,M1](http://books.google.com/books?id=LC3DfjWfCiwC&pg=PA239&lpg=PA239&dq=)). Cambridge University Press. pp. 238–241.
- Pinker, Steven (1994). *The Language Instinct: How the Mind Creates Language*. Perennial.
- Romaine, Suzanne (2001). "Multilingualism". In Mark Aronoff. *The Handbook of Linguistics*. Blackwell. pp. 512–533.
- Saussure, Ferdinand de (1983) [1913]. Bally, Charles; Sechehaye, Albert. eds. *Course in General Linguistics*. La Salle, Illinois: Open Court. ISBN 0-8126-9023-0.
- Sandler, Wendy; Lillo-Martin, Diane (2001). "Natural Sign Languages". In Mark Aronoff. *The Handbook of Linguistics*. Blackwell. pp. 533–563.
- Swadesh, Morris (1934). "The phonemic principle". *Language* **10** (2): 117–129. doi:10.2307/409603. JSTOR 409603.
- Tomasello, Michael (2008). *Origin of Human Communication*. MIT Press.
- Thomason, Sarah G.; Kaufman, Terrence (1988). *Language Contact, Creolization and Genetic Linguistics*. University of California Press.
- Thomason, Sarah G. (2001). *Language Contact - An Introduction*. Edinburgh University Press.
- Trask, Robert Lawrence (1999). *Language: The Basics* (2nd ed.). Psychology Press.
- Trask, Robert Lawrence (2007). Stockwell, Peter. ed. *Language and Linguistics: The Key Concepts* (2nd ed.). Routledge.
- Ulbaek, Ib (1998). "The Origin of Language and Cognition". In J. R. Hurford & C. Knight. *Approaches to the evolution of language*. Cambridge University Press. pp. 30–43.

Van Valin, jr, Robert D. (2001). "Functional Linguistics". In Mark Aronoff. *The Handbook of Linguistics*. Blackwell. pp. 319–337.

Zentella, Ana Celia (2002). "Spanish in New York". In García, Ofelia; Fishman, Joshua. *The Multilingual Apple: Languages in New York City*.

## External links

- World Atlas of Language Structures: a large database of structural (phonological, grammatical, lexical) properties of languages (<http://wals.info/>)

# Collective unconscious

**Collective unconscious** is a term of analytical psychology, coined by Carl Jung. It is proposed to be a part of the unconscious mind, expressed in humanity and all life forms with nervous systems, and describes how the structure of the psyche autonomously organizes experience. Jung distinguished the collective unconscious from the personal unconscious, in that the personal unconscious is a personal reservoir of experience unique to each individual, while the collective unconscious collects and organizes those personal experiences in a similar way with each member of a particular species.

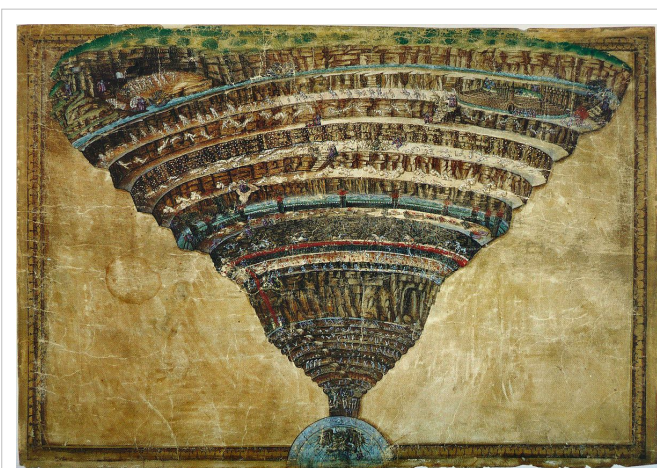


Illustration of the structure of Hell according to Dante Alighieri's *Divine Comedy*. By Sandro Botticelli (1480 ou 1495). According to Carl Gustav Jung, hell represents, among every culture, the disturbing aspect of the collective unconscious.

## Jung's definitions

For Jung, "My thesis then, is as follows: in addition to our immediate consciousness, which is of a thoroughly personal nature and which we believe to be the only empirical psyche (even if we tack on the personal unconscious as an appendix), there exists a second psychic system of a collective, universal, and impersonal nature which is identical in all individuals. This collective unconscious does not develop individually but is inherited. It consists of pre-existent forms, the archetypes, which can only become conscious secondarily and which give definite form to certain psychic contents."<sup>[1]</sup>

Jung linked the collective unconscious to 'what Freud called "archaic remnants" - mental forms whose presence cannot be explained by anything in the individual's own life and which seem to be aboriginal, innate, and inherited shapes of the human mind'.<sup>[2]</sup>



## Archetypes and collective representations

Jung considered that 'the shadow' and the anima/animus differ from the other archetypes in the fact that their content is more directly related to the individual's personal situation,<sup>[3]</sup> and less to the collective unconscious: by contrast, 'the collective unconscious is personified as a Wise Old Man'.<sup>[4]</sup>

Jung also made reference to contents of this category of the unconscious psyche as being similar to Levy-Bruhl's use of *collective representations* or "représentations collectives," Mythological "motifs," Hubert and Mauss's "categories of the imagination," and Adolf Bastian's "primordial thoughts."

## Minimal/maximal interpretations

In a minimalist interpretation of what would then appear as 'Jung's much misunderstood idea of the collective unconscious', his idea was 'simply that certain structures and predispositions of the unconscious are common to all of us...[on] an inherited, species-specific, genetic basis'.<sup>[5]</sup> Thus 'one could as easily speak of the "collective arm" - meaning the basic pattern of bones and muscles which all human arms share in common'.<sup>[6]</sup>

Others point out however that 'there does seem to be a basic ambiguity in Jung's various descriptions of the Collective Unconscious. Sometimes he seems to regard the predisposition to experience certain images as understandable in terms of some genetic model'<sup>[7]</sup> - as with the collective arm. However, Jung was 'also at pains to stress the numinous quality of these experiences, and there can be no doubt that he was attracted to the idea that the archetypes afford evidence of some communion with some divine or world mind', and perhaps 'his popularity as a thinker derives precisely from this'<sup>[8]</sup> - the maximal interpretation.

Marie-Louise von Franz accepted that 'it is naturally very tempting to identify the hypothesis of the collective unconscious historically and regressively with the ancient idea of an all-extensive world-soul'.<sup>[9]</sup> New Age writer Healy goes further, claiming that Jung himself 'dared to suggest that the human mind could link to ideas and motivations called the collective unconscious...a body of unconscious energy that lives forever'.<sup>[10]</sup> This is the idea of monopsychism.

## References

- [1] C. G. Jung, *The Archetypes and the Collective Unconscious* (London 1996) p. 43
- [2] C. G. Jung, *Man and his Symbols* (London 1978) p. 57
- [3] Walter A. Shelburne, *Mythos and Logos in the Thought of Carl Jung* (1988) p. 150
- [4] C. G. Jung, *Mysterium Coniunctionis* (London 1963) p. 106
- [5] Stan Gooch, *Total Man* (London 1975) p. 433
- [6] Gooch, p. 433
- [7] D. A G. Cook, "Jung" in Richard L. Gregory, *The Oxford Companion to the Mind* (Oxford 1987) p. 405
- [8] Cook, p. 405
- [9] Marie-Louise von Franz, *Projection and Re-Collection in Jungian Psychology* (1985) p. 85
- [10] Sherry Healy, *Dare to be Intuitive* (2005) p. 10

## Further reading

- Michael Vannoy Adams, *The Mythological Unconscious* (2001)
- Gallo, Ernest. "Synchronicity and the Archetypes," *Skeptical Inquirer*, 18 (4). Summer 1994.
- Jung, Carl. (1959). *Archetypes and the Collective Unconscious*.
- Jung, Carl. *The Development of Personality*.
- Jung, Carl. (1970). "Psychic conflicts in a child.", *Collected Works of C. G. Jung*, 17. Princeton University Press. 235 p. (p. 1-35).
- Whitmont, Edward C. (1969). *The Symbolic Quest*. Princeton University Press.

## External links

- for Research in Archetypal Symbolism (<http://aras.org>"Archive) A pictorial and written archive of mythological, ritualistic, and symbolic images from all over the world and from all epochs of human history.
- Kaleidoscope Forum (<http://kaleidoscope-forum.org/talk/index.php?sid=a17bf926549688279ac843e117322bd2>) Jungian Discussion Forum. All levels of discourse welcomed.
- Collective Unconscious at Carl Jung ([http://www.carl-jung.net/collective\\_unconscious.html](http://www.carl-jung.net/collective_unconscious.html))

## Jungian archetypes

---

Carl Gustav Jung developed an understanding of archetypes as being "ancient or archaic images that derive from the collective unconscious".<sup>[1]</sup> These are different from instincts, as Jung understood instincts as being "an unconscious physical impulse toward actions and the archetype as the psychic counterpart".<sup>[2]</sup> There are many different archetypes and Jung has stated they are limitless, but they have been simplified; examples include the persona, the shadow, the anima, the animus, the great mother, the wise old man, the hero, and the self.<sup>[2]</sup> The great mother, the wise old man, and the hero tend to be considered add-ons from the basic set, because they are not included in Jung's map of the soul along with the others. The archetypes can be used for a sense of understanding as well as for a state of treatment<sup>[2][3][4][5]</sup> "The archetype is a tendency to form such representations of a motif - representations that can vary a great deal in detail without losing their basic pattern ... They are indeed an instinctive *trend*".<sup>[6]</sup> Thus, "the archetype of initiation is strongly activated to provide a meaningful transition ... with a 'rite of passage' from one stage of life to the next".<sup>[7][8]</sup> such stages may include being parented, initiation, courtship, marriage and preparation for death.<sup>[9]</sup>

## Introduction

Virtually alone among the depth psychologists of the 20th century, Jung rejected the *tabula rasa* theory of human psychological development, believing instead that evolutionary pressures have individual predestinations manifested in archetypes. For Jung, "the archetype is the introspectively recognizable form of *a priori* psychic orderedness".<sup>[10]</sup> These images must be thought of as lacking in solid content, hence as unconscious. They only acquire solidity, influence, and eventual consciousness in the encounter with empirical facts."<sup>[11]</sup>

The archetypes form a dynamic substratum common to all humanity, upon the foundation of which each individual builds his own experience of life, developing a unique array of psychological characteristics. Thus, while archetypes themselves may be conceived as a relative few innate nebulous forms, from these may arise innumerable images, symbols and patterns of behavior. While the emerging images and forms are apprehended consciously, the archetypes which inform them are elementary structures which are unconscious and impossible to apprehend. Being unconscious, the existence of archetypes can only be deduced indirectly by examining behavior, images, art, myths, religions, dreams, etc. They are inherited potentials which are actualized when they enter consciousness as images or manifest in behavior on interaction with the outside world.

## Chronology

The intuition that there was more to the psyche than individual experience possibly began in Jung's childhood. The very first dream he could remember was that of an underground phallic god. His research on schizophrenia later supported his early intuition about the existence of universal psychic structures that underlie all human experience and behavior. Jung first referred to these as "primordial images" — a term he borrowed from Jacob Burckhardt. Later in 1917 Jung called them "dominants of the collective unconscious." It was not until 1919 that he first used the term "archetypes" in an essay titled "Instinct and the Unconscious". A main part of the chronology of Jung's discovery of the archetypes is found in the Redbook which documented Jung's encounter with the archetypes and

---

collective unconscious.<sup>[12]</sup>

## Origins

Jung began seeing patterns in his dreams and daily life in his early age. However, it was not until his later life, when he began piecing them together through archetypes, that he came to understand what these dreams actually meant.<sup>[13]</sup> These times were covered in the Red Book,<sup>[14]</sup> and the symbols that the archetypes represented and their origins in detail could be found in a Man and His Symbols. Jung suggested that the archetypes have always existed and will always exist as part of the collective unconscious.<sup>[15]</sup> It is sometimes assumed that people are creating new archetypes. However, they are not actually being created but rather discovered, and the number of archetypes in the world is limitless. Finding new archetypes is a matter of searching deep within one's self. The origins of the archetypal hypothesis date back to Plato. Jung himself compared archetypes to Platonic εἶδος (eidos): Plato's *ideas* which were pure mental forms imprinted in the soul before born into the world. They were collective in the sense that they embodied the fundamental characteristics of a thing rather than its specific characteristics. In fact, many of Jung's ideas were prevalent in Athenian philosophy. The archetype theory can be seen as a psychological equivalent to the philosophical idea of forms.

## Examples and conceptual difficulties

An archetype is a well recognized idea in psychology and many outside of psychology know the term as well, but many people find the topic or the idea behind the archetypes confusing. The confusion about the archetypes can partly be attributed to Jung's own evolving ideas about them in his writings and his interchangeable use of the term "archetype" and "primordial image"; it may also be attributed to the fact that, given his belief that "archetypal symbols ... are spontaneous and autonomous products of the unconscious", Jung was always intent "not to weaken the specific individual and cultural values of archetypes by leveling them out - i.e., by giving them a stereotyped, intellectually formulated meaning".<sup>[16]</sup>

Strictly speaking, archetypal figures such as the hero, the goddess and the wise man are not archetypes, but archetypal images which have crystallized out of the *archetypes-as-such*: as Jung put it, "definite mythological images of motifs ... are nothing more than conscious representations; it would be absurd to assume that such variable representations could be inherited", as opposed to their deeper, instinctual sources - "the 'archaic remnants', which I call 'archetypes' or 'primordial images'".<sup>[17]</sup>

Jung described archetypal events: birth, death, separation from parents, initiation, marriage, the union of opposites etc.; archetypal figures: great mother, father, child, devil, God, wise old man, wise old woman, Apollo, trickster,<sup>[18]</sup> hero - not to mention "Oedipus ... the first archetype Freud discovered"<sup>[19]</sup> or "number ... an *archetype of order*";<sup>[20]</sup> and archetypal motifs: the Apocalypse, the Deluge, the Creation, etc. Although the number of archetypes is limitless, there are a few particularly notable, recurring archetypal images, "the chief among them being" (according to Jung) "the *shadow*, the *Wise Old Man*, the *child* (including the child hero), the *mother* ... and her counterpart, the *maiden*, and lastly the *anima* in man and the *animus* in woman".<sup>[21]</sup> Alternatively he would speak of "the emergence of certain definite archetypes ... the shadow, the animal, the wise old man, the anima, the animus, the mother, the child".<sup>[22]</sup> Although five main archetypes were discussed in Jung's writing there are many others. The following are the five most common archetypes.<sup>[23]</sup>

Five main archetypes are sometimes enumerated:<sup>[24]</sup>

- The Self is the regulating center of the psyche and facilitator of individuation - the representative of "that wholeness which the introspective philosophy of all times and climes has characterized with an inexhaustible variety of symbols, names and concepts".<sup>[25]</sup> It represents all that is unique within a human being. Although a person is a collection of all the archetypes and what they learn from the collective unconscious, the self is what makes that person an I. The self cannot exist without the other archetypes and the other archetypes cannot exist

without the self; Jung makes this very clear. The self is also the part which grows and changes as a person goes throughout life.<sup>[23]</sup> The self can be summed up as the ideal form a person wishes to be.<sup>[26]</sup>

- The Shadow represents the traits which lie deep within ourselves. The traits that are hidden from day to day life and are in some cases the opposite of the self is a simple way to state these traits. The shadow is a very important trait because for one to truly know themselves, one must know all their traits, including those which lie beneath the common, i.e., the shadow. If one chooses to know the shadow there is a chance they give in to its motivation.<sup>[27]</sup>
- The Anima is sometimes seen -- e.g. by Campbell<sup>[28]</sup> -- as the feminine side within a man, but Jung did not fully intend this to be viewed in this way. The Anima is beyond generalization of society's views and stereotypes. Anima represents what femininity truly represents it in all its mysteries. It is what allows a man to be in touch with a woman.<sup>[23]</sup> The anima is commonly represented within dreams as a method to communicate with a person.<sup>[29]</sup> It contains all female encounters with men to help the relationship between the two improve better.
- The Animus is similar to the anima except for the fact that the animus allows a female to understand and communicate with a man.<sup>[24]</sup> Just like the anima, it is commonly represented in dreams of a woman to help them understand themselves and relationships with men<sup>[30]</sup> It can be known as part of the collective unconscious' connection with all of the encounters of males with females, like the anima, to improve relationship with males and females.
- The Persona is to Jung a mere "functional complex ... by no means identical to the individuality",<sup>[15]</sup> the way we present to the world - a mask which protects the Ego from negative images, and which by post-Jungians is sometimes considered an "archetype ... as a dynamic/structural component of the psyche".<sup>[16]</sup> Some view this as the opposite of the shadow which is not entirely true, this is just the face that is put on for the world, not our deepest internal secrets and desires; that is the self.<sup>[27]</sup>

However the precise relationships between images such as, for example, "the fish" and its archetype were not adequately explained by Jung. Here the image of the fish is not strictly speaking an archetype. However the "archetype of the fish" points to the ubiquitous existence of an innate "fish archetype" which gives rise to the fish image. In clarifying the contentious statement that fish archetypes are universal, Anthony Stevens explains that the *archetype-as-such* is at once an innate predisposition to form such an image and a preparation to encounter and respond appropriately to the creature per se. This would explain the existence of snake and spider phobias, for example, in people living in urban environments where they have never encountered either creature.<sup>[31]</sup> There are many examples such as the fish dealt with in *Man and His Symbols* and how they tend to relate to people through measures such as dreams and little life instances. These archetypal figures can also be represented from the main archetypes such as the anima and the animus or archetypal thoughts such as the resurrection of a savior figure<sup>[29]</sup> For example almost every culture has a savior that has come back from heaven or the dead, or reincarnation is a main point of the belief. Jesus for example in the Christian texts and also Buddhists and Hindus have reincarnation as a principal part of their religion, these being principal parts of many religions.

## Actualization and complexes

Archetypes seek actualization within the context of an individual's environment and determine the degree of individuation. Jung also used the terms "evocation" and "constellation" to explain the process of actualization. Thus for example, the mother archetype is actualized in the mind of the child by the evoking of innate anticipations of the maternal archetype when the child is in the proximity of a maternal figure who corresponds closely enough to its archetypal template. This mother archetype is built into the personal unconscious of the child as a mother *complex*. Complexes are functional units of the personal unconscious, in the same way that archetypes are units for the collective unconscious.

## Psychoid archetype

Jung proposed that the archetype had a dual nature: it exists both in the psyche and in the world at large. He called this non-psychic aspect of the archetype the "psychoid" archetype. He illustrated this by drawing on the analogy of the electromagnetic spectrum. The part of the spectrum which is visible to us corresponds to the conscious aspects of the archetype. The invisible infra-red end of the spectrum corresponds to the unconscious biological aspects of the archetype that merges with its chemical and physical conditions.<sup>[32]</sup> He suggested that not only do the archetypal structures govern the behavior of all living organisms, but that they were contiguous with structures controlling the behavior of inorganic matter as well. The archetype was not merely a psychic entity, but more fundamentally, a bridge to matter in general.<sup>[33]</sup> Jung used the term *unus mundus* to describe the unitary reality which he believed underlay all manifest phenomena. He conceived archetypes to be the mediators of the *unus mundus*, organizing not only ideas in the psyche, but also the fundamental principles of matter and energy in the physical world.

It was this psychoid aspect of the archetype that so impressed Nobel laureate physicist Wolfgang Pauli. Embracing Jung's concept, Pauli believed that the archetype provided a link between physical events and the mind of the scientist who studied them. In doing so he echoed the position adopted by German astronomer Johannes Kepler. Thus the archetypes which ordered our perceptions and ideas are themselves the product of an objective order which transcends both the human mind and the external world.<sup>[31]</sup>

## Parallels and developments

Although the term "archetype" did not originate with Jung, its current use has largely been influenced by his conception of it. The idea of innate psychic structures, at one time a relative novelty in the humanities and sciences has now been widely adopted.

### General developments

Related concepts arguably include the work of Claude Lévi-Strauss, an advocate of structuralism in anthropology, the concept of "social instincts" proposed by Charles Darwin, the "faculties" of Henri Bergson and the isomorphs of gestalt psychologist Wolfgang Kohler. In 1965 Noam Chomsky's ideas of human language acquisition being based on an "innate acquisition device" became known to the world.

Melanie Klein's idea of unconscious phantasy is closely related to Jung's archetype, as both are composed of image and affect and are *a priori* patternings of psyche whose contents are built from experience.

### Archetypal pedagogy

Archetypal pedagogy was developed by Clifford Mayes. Mayes' work also aims at promoting what he calls archetypal reflectivity in teachers; this is a means of encouraging teachers to examine and work with psychodynamic issues, images, and assumptions as those factors affect their pedagogical practices.

### Archetypes and psychology

Archetypal psychology was developed by James Hillman in the second half of the 20th century. Hillman trained at the Jung Institute and was its Director after graduation. Archetypal psychology is in the Jungian tradition and most directly related to analytical psychology and psychodynamic theory, yet departs radically. Archetypal psychology relativizes and deliteralizes the ego and focuses on the psyche, or soul, itself and the *archai*, the deepest patterns of psychic functioning, "the fundamental fantasies that animate all life".<sup>[34]</sup> Archetypal psychology is a polytheistic psychology, in that it attempts to recognize the myriad fantasies and myths gods, goddesses, demigods, mortals and animals—that shape and are shaped by our psychological lives. The ego is but one psychological fantasy within an assemblage of fantasies. The main influence on the development of archetypal psychology is Jung's analytical psychology. It is strongly influenced by Classical Greek, Renaissance, and Romantic ideas and thought. Influential

artists, poets, philosophers, alchemists, and psychologists include: Nietzsche, Henry Corbin, Keats, Shelley, Petrarch, and Paracelsus. Though all different in their theories and psychologies, they appear to be unified by their common concern for the psyche—the soul. Many archetypes have been used in treatment of psychological illnesses. Jung's first research was done with schizophrenics. A current example is teaching young men or boys archetypes through using picture books to help with the development.<sup>[35]</sup> In addition nurses treat patients through the use of archetypes.<sup>[7]</sup> Archetype therapy offers a wide range of uses if applied correctly, and it is still being expanded in Jungian schools today. With the list of archetypes being endless the healing possibilities are vast.

### **Robert Langs' use of archetypes**

Adaptive psychotherapist and psychoanalyst Robert Langs has recently begun using archetypal theory as a way of understanding the functioning of what he calls the "deep unconscious system".<sup>[36]</sup> Langs' use of archetypes particularly pertains to issues associated with death anxiety, which Langs takes to be the root of psychic conflict. Like Jung, Langs thinks of archetypes as species-wide, deep unconscious factors.<sup>[37]</sup>

### **Jung on the value of the archetype**

It is only possible to live the fullest life when we are in harmony with these symbols; wisdom is a return to them (CW8:794). To this it can be taken to mean that when a person is able to come to peace with the archetypes that lay within them they are able to begin to live a more peaceful life.

[For the alchemists] they were seeds of light broadcast in the chaos...the seed plot of a world to come...One would have to conclude from these alchemical visions that the archetypes have about them a certain effulgence, or quasi-consciousness, and that numinosity entails luminosity (CW8:388).

All the most powerful ideas in history go back to archetypes. This is particularly true of religious ideas, but the central concepts of science, philosophy, and ethics are no exception to this rule. In their present form they are variants of archetypal ideas created by consciously applying and adapting these ideas to reality. For it is the function of consciousness, not only to recognize and assimilate the external world through the gateway of the senses, but to translate into visible reality the world within us (CW8, 342). This could be taken to mean that the archetypes are what makes us, us. All of the beliefs and myths we have are all just part of the archetypes, nothing is new in the universe and everything has already existed and will continue to exist.

In his last text, *Man and His Symbols*, Jung stressed that "since so many people have chosen to treat archetypes as if they were part of a mechanical system that can be learned by rote, it is essential to insist that they are not mere names or even philosophical concepts. They are pieces of life itself - images that are integrally connected to the individual by the bridge of the emotions".<sup>[20]</sup> Jung states that they are not individual concepts of the world or individual pieces of the world we must come to know as separate things, but we must come to know the machine (archetypes) as a whole, not just as individuals.

As a result, it was the importance of the experiential encounter with the archetype which Jung emphasized: "in psychology, where we speak of archetypes like the anima and animus, the wise man, the great mother, and so on ... if they are mere images whose numinosity you have never experienced, it will be as if you were talking in a dream, for you will not know what you are talking about ... their names mean very little, whereas the way they are related to you is all-important".<sup>[21]</sup> This means that when one first encounters a new archetype, they do not always know what it means or how it will help in their life, but they must come to learn to accept it and understand it. In time, the answers will reveal themselves.

## Criticism of Jungian understandings

Lacan, in his "return to Freud", took issue with that aspect of "the thought of Jung, where the relation between the psychical world of the subject and reality are embodied under the term archetype".<sup>[38]</sup> He argued that "Jungianism - in so far as it makes of the primitive modes of articulating the world something that survives, the kernel, he says, of the psyche itself - is necessarily accompanied by a repudiation of the term *libido*".<sup>[39]</sup> Freud himself however had been well prepared to accept the existence of "a primitive kind of mental activity ... [on] the single analogy - and it is an excellent one - of the far-reaching *instinctive* knowledge of animals";<sup>[40]</sup> and it was indeed on the basis of "what Freud called 'archaic remnants' - mental forms whose presence cannot be explained by anything in the individual's own life ... inherited shapes of the human mind"<sup>[17]</sup> that Jung had explicitly built his theory of archetypes. His specific and contrasting claim was that they were "not in any sense lifeless or meaningless 'remnants'. They still function, and they are especially valuable ... just because of their 'historical' nature".<sup>[41]</sup>

More general criticism of the concept of archetypes can perhaps be placed in two broad categories. There are those who deny any possibility of inherited ideas as unscientific - a point met (at least to some degree) by Jung when he insisted that it was instead the inherited *propensity* to generate representations that made the archetypes "the unconscious organizers of our ideas"<sup>[42]</sup> (see above).

But those who could accept such inherited propensities still found "a basic ambiguity in Jung's various descriptions of the collective unconscious. Sometimes he seems to regard the predisposition to experience certain images as understandable in terms of some genetic model ... about the way human beings experience the world. But he is also at pains to emphasize the numinous quality of these experiences and there can be no doubt that he was attracted to the idea that the archetypes afford evidence of communion with some divine or world mind".<sup>[43]</sup> Jung's last statements on that subject remained however firmly agnostic. "Many people would agree with me if I stated flatly that such ideas are probably illusions ... [but] the denial is as impossible to 'prove' as the assertion".<sup>[44]</sup>

A more technical objection derives from therapeutic practice, with the possibility arising that "an explanation of the *archetypal* situation ... may lead to inflation, if it is not linked to specific and personal emotional experiences".<sup>[45]</sup> Some would go further, arguing that because "in Jungian theory, the psychologist's task is to lead others to see the timeless archetypal reality behind their personal psychological experiences ... using abstract, archetypal forces to explain human psychology", the result must inevitably be "a psychology which downplays the significance of human relationships".<sup>[46]</sup> The patient is thus brought to realise that "what I did then, what I felt then, is only the reflection of that great archetypal dream, or epic story ... free of the individual pain of it", but at the price of individuality and human relationship, sacrificed for an unwillingness to "leave the safety of myth".<sup>[47]</sup>

## References

- [1] Fiest J, Friest G, *Theories of Personality* (2009), New York New York; McGraw-Hill
- [2] Fiest J, Friest G, (2009) *Theories of Personality*, New York New York; McGraw-Hill
- [3] Jung, C. G. (1964), *Man and His Symbols*, Del Publishing, a division of Random House Inc.
- [4] *Jung's Map of the Soul an Introduction*, Peru Illinois: Carus Publishing Company
- [5] Aion Researches into the phenomenology of the self Princeton New Jersey: Princeton University Press
- [6] C. G. Jung, "Approaching the Unconscious" in C. G. Jung ed., *Man and his Symbols* (London 1978) p. 58
- [7] Rancour, P. (2008). Using archetypes and transitions theory to help patients move from active treatment to survivorship. *Clinical Journal of Oncology Nursing*, 12(6), 935-940. Retrieved from EBSCOhost.
- [8] Joseph Henderson, "Ancient Myths and Modern Man", in Jung ed., *Symbols* p. 123
- [9] Stevens, Anthony in "The Archetypes" (Chapter 3). Papadopoulos, Renos ed. (2006). *The Handbook of Jungian Psychology*.
- [10] C. G. Jung, *Synchronicity* (London 1985) p. 140
- [11] Jung 1928:Par. 300
- [12] Shamdassani, S., & Beebe, J. (2010). Jung Becomes Jung: A Dialogue on Liber Novus (The Red Book). *Psychological Perspectives*, 53(4), 410-436. doi:10.1080/00332925.2010.524110
- [13] *Theories of Personality* Fiest Jes, Fiest Gregory
- [14] 2 Shamdassani, S., & Beebe, J. (2010). Jung Becomes Jung: A Dialogue on Liber Novus (The Red Book). *Psychological Perspectives*, 53(4), 410-436. doi:10.1080/00332925.2010.524110

- [15] Man and His Symbols Jung Carl
- [16] M.-L. von Franz, "Science and the unconscious", in Jung ed., *Symbols* p. 386 and p. 377
- [17] Jung, "Approaching the Unconscious" in Jung ed., *Symbols* p. 57
- [18] den Uijl, S. (2010). The Trickster "Archetype" in the Shahnama. *Iranian Studies*, 43(1), 71-90. doi:10.1080/00210860903451220
- [19] Richard I. Evans, *Jung on Elementary Psychology* (London 1979) p. 52
- [20] Jung, *Synchronicity* p. 58
- [21] Jung, quoted in J. Jacobi, *Complex, Archetype, Symbol* (London 1959) p. 114
- [22] C. G. Jung, *Two Essays on Analytical Psychology* (London 1953) p. 108
- [23] Jung's Map of the Soul, Stein Muray.
- [24] *Jung's Map of the Soul* by Murray Stein.
- [25] C. G. Jung, *Alchemical Studies* (London 1978) p. 180-1
- [26] Aion Researches into the phenomenology of the self, Jung Carl
- [27] Theories of Personality, Fiest Jess, and Feist Gregory.
- [28] Joseph Campbell: "The Archetypes of the Collective Unconscious", in "Pathways to Bliss. Mythology and Personal Transformation" (David Kudler, ed.), 2004
- [29] Man and His symbols. Jung. Carl.
- [30] *Man and his Symbols*, C.G. Jung.
- [31] Stevens, Anthony in "The archetypes" (Chapter 3.) Ed. Papadopoulos, Renos. *The Handbook of Jungian Psychology* (2006)
- [32] Jung, C.G. (1947/1954) par. 420 *Collected Works*.
- [33] Jung, C.G. (1947/1954) par. 420 *Collected Works*
- [34] Moore, in Hillman, 1991
- [35] Zambo, D. (2007). Using Picture Books to Provide Archetypes to Young Boys: Extending the Ideas of William Brozo. *Reading Teacher*, 61(2), 124. Retrieved from EBSCOhost.t.
- [36] R Langs. *Fundamentals of Adaptive Psychotherapy and Counseling*. (London 2004)
- [37] R Langs. *Freud on a Precipice. How Freud's Fate pushed Psychoanalysis over the Edge*. (Lanham MD: 2010
- [38] Jacques Lacan, *The Four Fundamental Concepts of Psycho-Analysis* (London 1994) p. 152-3
- [39] Lacan, *Four* p. 153
- [40] Sigmund Freud, *Case Histories II* (London 1991) p. 364
- [41] Jung, "Approaching the Unconscious" in Jung ed., *Symbols* p. 32
- [42] C. G. Jung, *Aion* (London 1959) p. 179
- [43] David Cook, "Jung", in Richard Gregory ed., *The Oxford Companion to the Mind* (Oxford 1987) p. 405
- [44] Jung, "Approaching the Unconscious" in Jung ed., *Symbols* p. 76
- [45] Mario Jacoby, *The Analytic Encounter* (Canada 1984) p. 83
- [46] Naomi R. Goldbenberg, *Resurrecting the Body* (New York 1993) pp. 72, 130, and 98
- [47] Doris Lessing, *The Golden Notebook* (Herts 1973) p. 457-8

## Further reading

- Stevens, Anthony (2006), "Chapter 3", in Papadopoulos, Renos, *The Handbook of Jungian Psychology*
- Jung, C. G. (1917, 1928), *Two Essays on Analytical Psychology*, *Collected Works*, 7 (2 ed.), London: Routledge (published 1966)
- Jung, C. G. (1934–1954), *The Archetypes and The Collective Unconscious*, *Collected Works*, 9 (2 ed.), Princeton, NJ: Bollingen (published 1981), ISBN 0-691-01833-2



# Schumann resonances

The **Schumann resonances (SR)** are a set of spectrum peaks in the extremely low frequency (ELF) portion of the Earth's electromagnetic field spectrum. Schumann resonances are global electromagnetic resonances, excited by lightning discharges in the cavity formed by the Earth's surface and the ionosphere.

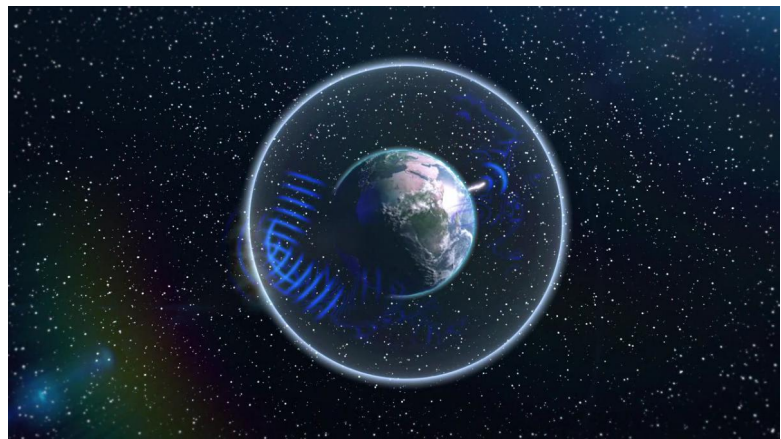
## Description

This global electromagnetic resonance phenomenon is named after physicist

Winfried Otto Schumann who predicted it mathematically in 1952. Schumann resonances occur because the space between the surface of the Earth and the conductive ionosphere acts as a closed waveguide. The limited dimensions of the Earth cause this waveguide to act as a resonant cavity for electromagnetic waves in the ELF band. The cavity is naturally excited by electric currents in lightning. Schumann resonances are the principal background in the electromagnetic spectrum<sup>[1]</sup> beginning at 3 Hz and extend to 60 Hz,<sup>[2]</sup> and appear as distinct peaks at extremely low frequencies (ELF) around 7.83 (fundamental),<sup>[3]</sup> 14.3, 20.8, 27.3 and 33.8 Hz.<sup>[4][5]</sup>

In the normal mode descriptions of Schumann resonances, the fundamental mode is a standing wave in the Earth–ionosphere cavity with a wavelength equal to the circumference of the Earth. This lowest-frequency (and highest-intensity) mode of the Schumann resonance occurs at a frequency of approximately 7.83 Hz, but this frequency can vary slightly from a variety of factors, such as solar-induced perturbations to the ionosphere, which comprises the upper wall of the closed cavity. The higher resonance modes are spaced at approximately 6.5 Hz intervals, a characteristic attributed to the atmosphere's spherical geometry. The peaks exhibit a spectral width of approximately 20% on account of the damping of the respective modes in the dissipative cavity. The 8th partial lies at approximately 60 Hz.

Observations of Schumann resonances have been used to track global lightning activity. Owing to the connection between lightning activity and the Earth's climate it has been suggested that they may also be used to monitor global temperature variations and variations of water vapor in the upper troposphere. It has been speculated that extraterrestrial lightning (on other planets) may also be detected and studied by means of their Schumann resonance signatures. Schumann resonances have been used to study the lower ionosphere on Earth and it has been suggested as one way to explore the lower ionosphere on celestial bodies. Effects on Schumann resonances have been reported following geomagnetic and ionospheric disturbances. More recently, discrete Schumann resonance excitations have been linked to transient luminous events – sprites, elves, jets, and other upper-atmospheric lightning. A new field of interest using Schumann resonances is related to short-term earthquake prediction.



Animation of Schumann resonance in Earth's atmosphere.

## History

The first documented observations of global electromagnetic resonance were made by Nikola Tesla at his Colorado Springs laboratory in 1899. This observation led to certain conclusions about the electrical properties of the Earth, and which made the basis for his idea for wireless energy transmission.<sup>[6]</sup>

Tesla researched ways to transmit power and energy wirelessly over long distances (via transverse waves and longitudinal waves). He transmitted extremely low frequencies through the ground as well as between the Earth's surface and the Kennelly-Heaviside layer. He received patents on wireless transceivers that developed standing waves by this method. Making mathematical calculations based on his experiments, Tesla discovered that the resonant frequency of the Earth was approximately 8 hertz (Hz). In the 1950s, researchers confirmed that the resonant frequency of the Earth's ionospheric cavity was in this range (later named the Schumann resonance).

In 1893, George Francis FitzGerald noted that the upper layers of the atmosphere must be fairly good conductors. Assuming that the height of these layers are about 100km above ground, he estimated that oscillations (in this case the lowest mode of the Schumann resonances) would have a period of 0.1 second.<sup>[7]</sup> Because of this contribution, it has been suggested to rename these resonances as Schumann–FitzGerald resonances.<sup>[8]</sup> However FitzGerald's findings were not widely known as they were only presented at a meeting of the British Association for the Advancement of Science, followed by a brief mention in a column in Nature.

Hence the first suggestion that an ionosphere existed, capable of trapping electromagnetic waves, is attributed to Heaviside and Kennelly (1902).<sup>[9][10]</sup> It took another twenty years before Edward Appleton and Barnett in 1925,<sup>[11]</sup> were able to prove experimentally the existence of the ionosphere.

Although some of the most important mathematical tools for dealing with spherical waveguides were developed by G. N. Watson in 1918,<sup>[12]</sup> it was Winfried Otto Schumann who first studied the theoretical aspects of the global resonances of the earth–ionosphere waveguide system, known today as the Schumann resonances. In 1952–1954 Schumann, together with H. L. König, attempted to measure the resonant frequencies.<sup>[13][14][15][16]</sup> However, it was not until measurements made by Balsler and Wagner in 1960–1963<sup>[17][18][19][20][21]</sup> that adequate analysis techniques were available to extract the resonance information from the background noise. Since then there has been an increasing interest in Schumann resonances in a wide variety of fields.

## Basic theory

Lightning discharges are considered to be the primary natural source of Schumann resonance excitation; lightning channels behave like huge antennas that radiate electromagnetic energy at frequencies below about 100 kHz.<sup>[22]</sup> These signals are very weak at large distances from the lightning source, but the Earth–ionosphere waveguide behaves like a resonator at ELF frequencies and amplifies the spectral signals from lightning at the resonance frequencies.<sup>[22]</sup>

In an ideal cavity, the resonant frequency of the  $n$ -th mode  $f_n$  is determined by the Earth radius  $a$  and the speed of light  $c$ .<sup>[13]</sup>

$$f_n = \frac{c}{2\pi a} \sqrt{n(n+1)}$$

The real Earth–ionosphere waveguide is not a perfect electromagnetic resonant cavity. Losses due to finite ionosphere electrical conductivity lower the propagation speed of electromagnetic signals in the cavity, resulting in a resonance frequency that is lower than would be expected in an ideal case, and the observed peaks are wide. In addition, there are a number of horizontal asymmetries – day-night difference in the height of the ionosphere, latitudinal changes in the Earth magnetic field, sudden ionospheric disturbances, polar cap absorption, variation in the Earth radius of +/- 11km from equator to geographic poles, etc. that produce other effects in the Schumann resonance power spectra.

## Measurements

Today Schumann resonances are recorded at many separate research stations around the world. The sensors used to measure Schumann resonances typically consist of two horizontal magnetic inductive coils for measuring the north-south and east-west components of the magnetic field, and a vertical electric dipole antenna for measuring the vertical component of the electric field. A typical passband of the instruments is 3–100 Hz. The Schumann resonance electric field amplitude (~300 microvolts per meter) is much smaller than the static fair-weather electric field (~150 V/m) in the atmosphere. Similarly, the amplitude of the Schumann resonance magnetic field (~1 picotesla) is many orders of magnitude smaller than the Earth magnetic field (~30–50 microteslas).<sup>[23]</sup> Specialized receivers and antennas are needed to detect and record Schumann resonances. The electric component is commonly measured with a ball antenna, suggested by Ogawa et al., in 1966,<sup>[24]</sup> connected to a high-impedance amplifier. The magnetic induction coils typically consist of tens- to hundreds-of-thousands of turns of wire wound around a core of very high magnetic permeability.

## Dependence on global lightning activity

From the very beginning of Schumann resonance studies, it was known that they could be used to monitor global lightning activity. At any given time there are about 2000 thunderstorms around the globe.<sup>[25]</sup> Producing ~50 lightning events per second,<sup>[26]</sup> these thunderstorms create the background Schumann resonance signal.

Determining the spatial lightning distribution from Schumann resonance records is a complex problem: in order to estimate the lightning intensity from Schumann resonance records it is necessary to account for both the distance to lightning sources as well as the wave propagation between the source and the observer. The common approach is to make a preliminary assumption on the spatial lightning distribution, based on the known properties of lightning climatology. An alternative approach is placing the receiver at the North or South Pole, which remain approximately equidistant from the main thunderstorm centers during the day.<sup>[27]</sup> One method not requiring preliminary assumptions on the lightning distribution<sup>[28]</sup> is based on the decomposition of the average background Schumann resonance spectra, utilizing ratios between the average electric and magnetic spectra and between their linear combination. This technique assumes the cavity is spherically symmetric and therefore does not include known cavity asymmetries that are believed to affect the resonance and propagation properties of electromagnetic waves in the system.

## Diurnal variations

The best documented and the most debated features of the Schumann resonance phenomenon are the diurnal variations of the background Schumann resonance power spectrum.

A characteristic Schumann resonance diurnal record reflects the properties of both global lightning activity and the state of the Earth–ionosphere cavity between the source region and the observer. The vertical electric field is independent of the direction of the source relative to the observer, and is therefore a measure of global lightning. The diurnal behavior of the vertical electric field shows three distinct maxima, associated with the three "hot spots" of planetary lightning activity: 9 UT (Universal Time) peak, linked to the increased thunderstorm activity from south-east Asia; 14 UT peak associated with the peak in African lightning activity; and the 20 UT peak resulting from the increase in lightning activity in South America. The time and amplitude of the peaks vary throughout the year, reflecting the seasonal changes in lightning activity.

### "Chimney" ranking

In general, the African peak is the strongest, reflecting the major contribution of the African "chimney" to the global lightning activity. The ranking of the two other peaks – Asian and American – is the subject of a vigorous dispute among Schumann resonance scientists. Schumann resonance observations made from Europe show a greater contribution from Asia than from South America. This contradicts optical satellite and climatological lightning data that show the South American thunderstorm center stronger than the Asian center,<sup>[26]</sup> although observations made from North America indicate the dominant contribution comes from South America. The reason for such disparity remains unclear, but may have something to do with the 60 Hz cycling of electricity used in North America (60 Hz being a mode of Schumann Resonance). Williams and Satori<sup>[29]</sup> suggest that in order to obtain "correct" Asia-America chimney ranking, it is necessary to remove the influence of the day/night variations in the ionospheric conductivity (day-night asymmetry influence) from the Schumann resonance records. On the other hand, such "corrected" records presented in the work by Satori et al.<sup>[30]</sup> show that even after the removal of the day-night asymmetry influence from Schumann resonance records, the Asian contribution remains greater than American. Similar results were obtained by Pechony et al.<sup>[31]</sup> who calculated Schumann resonance fields from satellite lightning data. It was assumed that the distribution of lightning in the satellite maps was a good proxy for Schumann excitations sources, even though satellite observations predominantly measure in-cloud lightning rather than the cloud-to-ground lightning that are the primary excitors of the resonances. Both simulations – those neglecting the day-night asymmetry, and those taking this asymmetry into account, showed same Asia-America chimney ranking. As for today, the reason for the "invert" ranking of Asia and America chimneys in Schumann resonance records remains unclear and the subject requires further, targeted research.

### Influence of the day-night asymmetry

In the early literature the observed diurnal variations of Schumann resonance power were explained by the variations in the source-receiver (lightning-observer) geometry.<sup>[17]</sup> It was concluded that no particular systematic variations of the ionosphere (which serves as the upper waveguide boundary) are needed to explain these variations.<sup>[32]</sup> Subsequent theoretical studies supported the early estimations of the small influence of the ionosphere day-night asymmetry (difference between day-side and night-side ionosphere conductivity) on the observed variations in Schumann resonance field intensities.<sup>[33]</sup>

The interest in the influence of the day-night asymmetry in the ionosphere conductivity on Schumann resonances gained new strength in the 1990s, after publication of a work by Sentman and Fraser.<sup>[34]</sup> Sentman and Fraser developed a technique to separate the global and the local contributions to the observed field power variations using records obtained simultaneously at two stations that were widely separated in longitude. They interpreted the diurnal variations observed at each station in terms of a combination of a diurnally varying global excitation modulated by the local ionosphere height. Their work, which combined both observations and energy conservation arguments, convinced many scientists of the importance of the ionospheric day-night asymmetry and inspired numerous experimental studies. However, recently it was shown that results obtained by Sentman and Fraser can be approximately simulated with a uniform model (without taking into account ionosphere day-night variation) and therefore cannot be uniquely interpreted solely in terms of ionosphere height variation.<sup>[35]</sup>

Schumann resonance amplitude records show significant diurnal and seasonal variations which in general coincide in time with the times of the day-night transition (the terminator). This time-matching seems to support the suggestion of a significant influence of the day-night ionosphere asymmetry on Schumann resonance amplitudes. There are records showing almost clock-like accuracy of the diurnal amplitude changes.<sup>[30]</sup> On the other hand there are numerous days when Schumann Resonance amplitudes do not increase at sunrise or do not decrease at sunset. There are studies showing that the general behavior of Schumann resonance amplitude records can be recreated from diurnal and seasonal thunderstorm migration, without invoking ionospheric variations.<sup>[31][33]</sup> Two recent independent theoretical studies have shown that the variations in Schumann resonance power related to the day-night transition are much smaller than those associated with the peaks of the global lightning activity, and therefore the

global lightning activity plays a more important role in the variation of the Schumann resonance power.<sup>[31][36]</sup>

It is generally acknowledged that source-observer effects are the dominant source of the observed diurnal variations, but there remains considerable controversy about the degree to which day-night signatures are present in the data. Part of this controversy stems from the fact that the Schumann resonance parameters extractable from observations provide only a limited amount of information about the coupled lightning source-ionospheric system geometry. The problem of inverting observations to simultaneously infer both the lightning source function and ionospheric structure is therefore extremely underdetermined, leading to the possibility of nonunique interpretations.

### **The "inverse problem"**

One of the interesting problems in Schumann resonances studies is determining the lightning source characteristics (the "inverse problem"). Temporally resolving each individual flash is impossible because the mean rate of excitation by lightning, ~50 lightning events per second globally, mixes up the individual contributions together. However, occasionally there occur extremely large lightning flashes which produce distinctive signatures that stand out from the background signals. Called "Q-bursts", they are produced by intense lightning strikes that transfer large amounts of charge from clouds to the ground, and often carry high peak current.<sup>[24]</sup> Q-bursts can exceed the amplitude of the background signal level by a factor of 10 or more, and appear with intervals of ~10 s,<sup>[28]</sup> which allows to consider them as isolated events and determine the source lightning location. The source location is determined with either multi-station or single-station techniques, and requires assuming a model for the Earth-ionosphere cavity. The multi-station techniques are more accurate, but require more complicated and expensive facilities.

### **Transient luminous events research**

It is now believed that many of the Schumann resonances transients (Q bursts) are related to the transient luminous events (TLEs). In 1995 Boccippio et al.<sup>[37]</sup> showed that sprites, the most common TLE, are produced by positive cloud-to-ground lightning occurring in the stratiform region of a thunderstorm system, and are accompanied by Q-burst in the Schumann resonances band. Recent observations<sup>[37][38]</sup> reveal that occurrences of sprites and Q bursts are highly correlated and Schumann resonances data can possibly be used to estimate the global occurrence rate of sprites.<sup>[39]</sup>

### **Global temperature**

Williams [1992]<sup>[40]</sup> suggested that global temperature may be monitored with the Schumann resonances. The link between Schumann resonance and temperature is lightning flash rate, which increases nonlinearly with temperature.<sup>[40]</sup> The nonlinearity of the lightning-to-temperature relation provides a natural amplifier of the temperature changes and makes Schumann resonance a sensitive "thermometer". Moreover, the ice particles that are believed to participate in the electrification processes which result in a lightning discharge<sup>[41]</sup> have an important role in the radiative feedback effects that influence the atmosphere temperature. Schumann resonances may therefore help us to understand these feedback effects. A strong link between global lightning and global temperature has not been experimentally confirmed as of 2008.

### **Upper tropospheric water vapor**

Tropospheric water vapor is a key element of the Earth's climate, which has direct effects as a greenhouse gas, as well as indirect effect through interaction with clouds, aerosols and tropospheric chemistry. Upper tropospheric water vapor (UTWV) has a much greater impact on the greenhouse effect than water vapor in the lower atmosphere,<sup>[42]</sup> but whether this impact is a positive, or a negative feedback is still uncertain.<sup>[43]</sup> The main challenge in addressing this question is the difficulty in monitoring UTWV globally over long timescales. Continental deep-convective thunderstorms produce most of the lightning discharges on Earth. In addition, they transport large amount of water vapor into the upper troposphere, dominating the variations of global UTWV. Price [2000]<sup>[44]</sup> suggested that changes in the UTWV can be derived from records of Schumann Resonances. According to the

effective work made by the Upper Tropospheric Water Vapor (( UTWV )), we should highlight that the percentage of UTWV in normal condition of the Air mass can be measured as a minimal quantity, so that its influence can be considered very very low; in fact the higher percentage of it can be only found in the lower Tropospheric layers. But in the case of a high quantity of UTWV in the highest level of Troposphere, due to a warmer air mass of Atlantic origins, for instance, the Water vapor, due to the low air temperature ((about minus 60 Degrees )) it turns into ice crystal, becoming clouds as Cirrus or Cirrus Stratus: no Water vapour exists as gas with so low temperature. So, we can say that the affirmation that Water vapor interacts with cloud, can be considered wrong as the clouds both those of low level of ((Atmosphere)) and those of higher levels of it are made of condensed or crystallised Water Vapor.

### **Schumann resonances on other planets**

The existence of Schumann-like resonances is conditioned primarily by two factors: (1) a closed, planetary-sized spherical cavity, consisting of conducting lower and upper boundaries separated by an insulating medium. For the earth the conducting lower boundary is its surface, and the upper boundary is the ionosphere. Other planets may have similar electrical conductivity geometry, so it is speculated that they should possess similar resonant behavior. (2) source of electrical excitation of electromagnetic waves in the ELF range. Within the Solar System there are five candidates for Schumann resonance detection besides the Earth: Venus, Mars, Jupiter, Saturn and its moon Titan.

Modeling Schumann resonances on the planets and moons of the Solar System is complicated by the lack of knowledge of the waveguide parameters. No in situ capability exists today to validate the results, but in the case of Mars there have been terrestrial observations of radio emission spectra that have been associated with Schumann resonances.<sup>[45]</sup> The reported radio emissions are not of the primary electromagnetic Schumann modes, but rather of secondary modulations of the nonthermal microwave emissions from the planet at approximately the expected Schumann frequencies, and have not been independently confirmed to be associated with lightning activity on Mars. There is the possibility that future lander missions could carry in situ instrumentation to perform the necessary measurements. Theoretical studies are primarily directed to parameterizing the problem for future planetary explorers.

The strongest evidence for lightning on Venus comes from the impulsive electromagnetic waves detected by Venera 11 and 12 landers. Theoretical calculations of the Schumann resonances at Venus were reported by Nickolaenko and Rabinowicz [1982]<sup>[46]</sup> and Pechony and Price [2004].<sup>[47]</sup> Both studies yielded very close results, indicating that Schumann resonances should be easily detectable on that planet given a lightning source of excitation and a suitably located sensor.

On Mars detection of lightning activity has been reported by Ruf et al. [2009].<sup>[45]</sup> The evidence is indirect and in the form of modulations of the nonthermal microwave spectrum at approximately the expected Schumann resonance frequencies. It has not been independently confirmed that these are associated with electrical discharges on Mars. In the event confirmation is made by direct, in situ observations, it would verify the suggestion of the possibility of charge separation and lightning strokes in the Martian dust storms made by Eden and Vonnegut [1973]<sup>[48]</sup> and Renno et al. [2003].<sup>[49]</sup> Martian global resonances were modeled by Sukhorukov [1991],<sup>[50]</sup> Pechony and Price [2004]<sup>[47]</sup> and Molina-Cuberos et al. [2006].<sup>[51]</sup> The results of the three studies are somewhat different, but it seems that at least the first two Schumann resonance modes should be detectable. Evidence of the first three Schumann resonance modes is present in the spectra of radio emission from the lightning detected in Martian dust storms.<sup>[45]</sup>

It was long ago suggested that lightning discharges may occur on Titan,<sup>[52]</sup> but recent data from Cassini–Huygens seems to indicate that there is no lightning activity on this largest satellite of Saturn. Due to the recent interest in Titan, associated with the Cassini–Huygens mission, its ionosphere is perhaps the most thoroughly modeled today. Schumann resonances on Titan have received more attention than on any other celestial body, in works by Besser et al. [2002],<sup>[53]</sup> Morente et al. [2003],<sup>[54]</sup> Molina-Cuberos et al. [2004],<sup>[55]</sup> Nickolaenko et al. [2003]<sup>[56]</sup> and Pechony and Price [2004].<sup>[47]</sup> It appears that only the first Schumann resonance mode might be detectable on Titan.

Jupiter is one planet where lightning activity has been optically detected. Existence of lightning activity on that planet was predicted by Bar-Nun [1975]<sup>[57]</sup> and it is now supported by data from Galileo, Voyagers 1 and 2, Pioneers 10 and 11 and Cassini. Saturn is also confirmed to have lightning activity. ([http://www.ciclops.org/view\\_event/178/Lightning\\_Flashing\\_in\\_Daylight](http://www.ciclops.org/view_event/178/Lightning_Flashing_in_Daylight)) Though three visiting spacecrafts – Pioneer 11 in 1979, Voyager 1 in 1980 and Voyager 2 in 1981, failed to provide any convincing evidence from optical observations, in July of 2012 the Cassini spacecraft detected visible lightning flashes, and electromagnetic sensors aboard the spacecraft detected signatures that are characteristic of lightning. Little is known about the electrical parameters of Jupiter and Saturn interior. Even the question of what should serve as the lower waveguide boundary is a non-trivial one in case of the gaseous planets. There seem to be no works dedicated to Schumann resonances on Saturn. To date there has been only one attempt to model Schumann resonances on Jupiter.<sup>[58]</sup> Here, the electrical conductivity profile within the gaseous atmosphere of Jupiter was calculated using methods similar to those used to model stellar interiors, and it was pointed out that the same methods could be easily extended to the other gas giants Saturn, Uranus and Neptune. Given the intense lightning activity at Jupiter, the Schumann resonances should be easily detectable with a sensor suitably positioned within the planetary-ionospheric cavity.

## References

- [1] The electrical nature of storms By D. R. MacGorman, W. D. Rust, W. David Rust. Page 114 ([http://books.google.com/books?id=\\_NbHNj7KJecC&pg=PA114](http://books.google.com/books?id=_NbHNj7KJecC&pg=PA114)).
- [2] Handbook of atmospheric electrodynamics, Volume 1 By Hans Volland. Page 277.
- [3] Can Resonant Oscillations of the Earth Ionosphere Influence the Human Brain Biorhythm? - V.D. Rusov, Department of Theoretical and Experimental Nuclear Physics, Odessa National Polytechnic University, Ukraine! url=[www.arxiv.org/pdf/1208.4970](http://www.arxiv.org/pdf/1208.4970)
- [4] The electrical nature of storms By D. R. MacGorman, W. D. Rust, W. David Rust. Page 114.
- [5] Recent advances in multidisciplinary applied physics By A. Méndez-Vilas. Page 65.
- [6] N. Tesla (1905). "The Transmission of Electrical Energy Without Wires As A Means Of Furthering World Peace". *Electrical World and Engineer* **January 7**: 21–24.
- [7] G. F. FitzGerald, "On the period of vibration of electrical disturbances upon the Earth," Br. Assoc. Adv. Sci., Rep. 63, 682 (1893)
- [8] J. D. Jackson, Examples of the zeroth theorem of the history of science, American Journal of Physics, Vol. 76, No. 8, pp. 704–719, August 2008
- [9] O. Heaviside (1902). "Telegraphy, Sect. 1, Theory". *Encyc. Brit. 10th ed.* . London **9**: 213–218.
- [10] A.E. Kennelly (1902). "On the elevation of the electrically-conducting strata of the earth's atmosphere". *Electrical world and engineer* **32**: 473–473.
- [11] Appleton, E. V. , M. A. F. Barnett (1925). "On Some Direct Evidence for Downward Atmospheric Reflection of Electric Rays". *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **109** (752): 621–641. Bibcode 1925RSPSA.109..621A. doi:10.1098/rspa.1925.0149.
- [12] Watson, G.N. (1918). "The diffraction of electric waves by the Earth". *Proc. Roy. Soc. (London)* **Ser.A 95**: 83–99.
- [13] Schumann W. O. (1952). "Über die strahlungslosen Eigenschwingungen einer leitenden Kugel, die von einer Luftschicht und einer Ionosphärenhülle umgeben ist". *Zeitschrift und Naturforschung* **7a**: 149–154. Bibcode 1952ZNatA...7..149S.
- [14] Schumann W. O. (1952). "Über die Dämpfung der elektromagnetischen Eigenschwingungen des Systems Erde – Luft – Ionosphäre". *Zeitschrift und Naturforschung* **7a**: 250–252. Bibcode 1952ZNatA...7..250S.
- [15] Schumann W. O. (1952). "Über die Ausbreitung sehr Langer elektrischer Wellen um die Signale des Blitzes". *Nuovo Cimento* **9** (12): 1116–1138. doi:10.1007/BF02782924.
- [16] Schumann W. O. and H. König (1954). "Über die Beobachtung von Atmospheric bei geringsten Frequenzen". *Naturwiss* **41** (8): 183–184. Bibcode 1954NW.....41..183S. doi:10.1007/BF00638174.
- [17] Balser M. and C. Wagner (1960). "Measurement of the spectrum of radio noise from 50 to 100 c/s". *J.Res. NBS* **64D**: 415–418.
- [18] Balser M. and C. Wagner (1960). "Observations of Earth–ionosphere cavity resonances". *Nature* **188** (4751): 638–641. Bibcode 1960Natur.188..638B. doi:10.1038/188638a0.
- [19] Balser M. and C. Wagner (1962). "Diurnal power variations of the Earth–ionosphere cavity modes and their relationship to worldwide thunderstorm activity". *J.G.R* **67** (2): 619–625. Bibcode 1962JGR....67..619B. doi:10.1029/JZ067i002p00619.
- [20] Balser M. and C. Wagner (1962). "On frequency variations of the Earth–ionosphere cavity modes". *J.G.R* **67** (10): 4081–4083. Bibcode 1962JGR....67.4081B. doi:10.1029/JZ067i010p04081.
- [21] Balser M. and C. Wagner (1963). "Effect of a high-altitude nuclear detonation on the Earth–ionosphere cavity". *J.G.R* **68**: 4115–4118.
- [22] Volland, H. (1984). *Atmospheric Electrodynamics*. Springer-Verlag, Berlin.
- [23] Price, C., O. Pechony, E. Greenberg (2006). "Schumann resonances in lightning research". *Journal of Lightning Research* **1**: 1– 15.
- [24] Ogawa, T., Y. Tanka, T. Miura, and M. Yasuhara (1966). "Observations of natural ELF electromagnetic noises by using the ball antennas". *J. Geomagn. Geoelectr* **18**: 443– 454.

- [25] Heckman S. J., E. Williams, (1998). "Total global lightning inferred from Schumann resonance measurements". *J. G. R.* **103(D24)**: 31775–31779. Bibcode 1998JGR...10331775H. doi:10.1029/98JD02648.
- [26] Christian H. J., R.J. Blakeslee, D.J. Boccippio, W.L. Boeck, D.E. Buechler, K.T. Driscoll, S.J. Goodman, J.M. Hall, W.J. Koshak, D.M. Mach, M.F. Stewart, (2003). "Global frequency and distribution of lightning as observed from space by the Optical Transient Detector". *J. G. R.* **108(D1)**: 4005. Bibcode 2003JGRD..108.4005C. doi:10.1029/2002JD002347.
- [27] Nickolaenko, A.P. (1997). "Modern aspects of Schumann resonance studies". *J.a.s.t.p.* **59**: 806–816.
- [28] Shvets A.V. (2001). "A technique for reconstruction of global lightning distance profile from background Schumann resonance signal". *J.a.s.t.p.* **63**: 1061–1074.
- [29] Williams E. R., G. Satori (2004). "Lightning, thermodynamic and hydrological comparison of the two tropical continental chimneys". *J.a.s.t.p.* **66**: 1213–1231.
- [30] Satori G., M. Neska, E. Williams, J. Szendrői (2007). "Signatures of the non-uniform Earth–ionosphere cavity in high time-resolution Schumann resonance records". *Radio Science* **in print**.
- [31] Pechony, O., C. Price, A.P. Nickolaenko (2007). "Relative importance of the day-night asymmetry in Schumann resonance amplitude records". *Radio Science* **in print**.
- [32] Madden T., W. Thompson (1965). "Low-frequency electromagnetic oscillations of the Earth–ionosphere cavity". *Rev. Geophys.* **3** (2): 211. Bibcode 1965RvGSP...3...211M. doi:10.1029/RG003i002p00211.
- [33] Nickolaenko A. P. and M. Hayakawa (2002). *Resonances in the Earth–ionosphere cavity*. Kluwer Academic Publishers, Dordrecht-Boston-London.
- [34] Sentman, D.D., B. J. Fraser (1991). "Simultaneous observations of Schumann Resonances in California and Australia – Evidence for intensity modulation by the local height of the D region". *Journal of geophysical research* **96** (9): 15973–15984. Bibcode 1991JGR...9615973S. doi:10.1029/91JA01085.
- [35] Pechony, O., C. Price (2006). "Schumann Resonances: interpretation of local diurnal intensity modulations". *Radio Sci.* **42** (2): RS2S05. Bibcode 2006RaSc...41.2S05P. doi:10.1029/2006RS003455.
- [36] Yang H., V. P. Pasko (2007). "Three-dimensional finite difference time domain modeling of the diurnal and seasonal variations in Schumann resonance parameters". *Radio Science* **41** (2): RS2S14. Bibcode 2006RaSc...41.2S14Y. doi:10.1029/2005RS003402.
- [37] Boccippio, D. J., E. R. Williams, S. J. Heckman, W. A. Lyons, I. T. Baker, R. Boldi (1995). "Sprites, ELF transients, and positive ground strokes". *Science* **269** (5227): 1088–1091. Bibcode 1995Sci...269.1088B. doi:10.1126/science.269.5227.1088. PMID 17755531.
- [38] Price, C., E. Greenberg, Y. Yair, G. Satori, J. Bór, H. Fukunishi, M. Sato, P. Israelevich, M. Moalem, A. Devir, Z. Levin, J.H. Joseph, I. Mayo, B. Ziv, A. Sternlieb (2004). "Ground-based detection of TLE-producing intense lightning during the MEIDEX mission on board the Space Shuttle Columbia". *G.R.L.* **31**.
- [39] Hu, W., S. A. Cummer, W. A. Lyons, T. E. Nelson (2002). "Lightning charge moment changes for the initiation of sprites". *G.R.L.* **29** (8): 1279. Bibcode 2002GeoRL..29h.120H. doi:10.1029/2001GL014593.
- [40] Williams, E.R. (1992). "The Schumann resonance: a global tropical thermometer". *Science* **256** (5060): 1184–1186. Bibcode 1992Sci...256.1184W. doi:10.1126/science.256.5060.1184. PMID 17795213.
- [41] Williams, E.R. (1989). "The tripole structure of thunderstorms". *J. G. R.* **94**: 13151–13167. Bibcode 1989JGR...9413151W. doi:10.1029/JD094iD11p13151.
- [42] Hansen, J., A. Lacis, D. Rind, G. Russel, P. Stone, I. Fung, R. Ruedy, J., Lerner (1984). "Climate sensitivity: Analysis of feedback mechanisms". *Climate Processes and Climate Sensitivity*, J., E. Hansen and T. Takahashi, eds.. *AGU Geophys. Monograph* **29**: 130–163.
- [43] Rind, D. (1998). "Just add water vapor". *Science* **28** (5380): 1152–1153. doi:10.1126/science.281.5380.1152.
- [44] Price, C. (2000). "Evidence for a link between global lightning activity and upper tropospheric water vapor". *Letters to Nature* **406** (6793): 290–293. doi:10.1038/35018543. PMID 10917527.
- [45] Ruf, C., N. O. Renno, J. F. Kok, E. Bandelier, M. J. Sander, S. Gross, L. Skjerve, and B. Cantor (2009). "Emission of Non-thermal Microwave Radiation by a Martian Dust Storm". *Geophys. Res. Lett.* **36** (13): L13202. Bibcode 2009GeoRL...3613202R. doi:10.1029/2009GL038715.
- [46] Nickolaenko A. P., L. M. Rabinowicz (1982). "On the possibility of existence of global electromagnetic resonances on the planets of Solar system". *Space Res.* **20**: 82–89.
- [47] Pechony, O., C. Price (2004). "Schumann resonance parameters calculated with a partially uniform knee model on Earth, Venus, Mars, and Titan". *Radio Sci.* **39** (5): RS5007. Bibcode 2004RaSc...39.5007P. doi:10.1029/2004RS003056.
- [48] Eden, H. F. and B. Vonnegut (1973). "Electrical breakdown caused by dust motion in low-pressure atmospheres: consideration for Mars". *Science* **180** (4089): 962–3. Bibcode 1973Sci...180..962E. doi:10.1126/science.180.4089.962. PMID 17735929.
- [49] Renno N. O., A. Wong, S. K. Atreya, I. de Pater, M. Roos-Serote (2003). "Electrical discharges and broadband radio emission by Martian dust devils and dust storms". *G. R. L.* **30** (22): 2140. Bibcode 2003GeoRL...30vPLA1R. doi:10.1029/2003GL017879.
- [50] Sukhorukov A. I. (1991). "On the Schumann resonances on Mars". *Planet. Space Sci.* **39** (12): 1673–1676. Bibcode 1991P&SS...39.1673S. doi:10.1016/0032-0633(91)90028-9.
- [51] Molina-Cuberos G. J., J. A. Morente, B. P. Besser, J. Porti, H. Lichtenegger, K. Schwingenschuh, A. Salinas, J. Margineda (2006). "Schumann resonances as a tool to study the lower ionosphere of Mars". *Radio Science* **41**: RS1003. Bibcode 2006RaSc...41.1003M. doi:10.1029/2004RS003187.
- [52] Lammer H., T. Tokano, G. Fischer, W. Stumptner, G. J. Molina-Cuberos, K. Schwingenschuh, H. O. Rucher (2001). "Lightning activity of Titan: can Cassini/Huygens detect it?". *Planet. Space Sci.* **49** (6): 561–574. Bibcode 2001P&SS...49..561L.



- doi:10.1016/S0032-0633(00)00171-9.
- [53] Besser, B. P., K. Schwingenschuh, I. Jernej, H. U. Eichelberger, H. I. M. Lichtenegger, M. Fulchignoni, G. J. Molina-Cuberos, J. A. Morente, J. A. Porti, A. Salinas (2002). "Schumann resonances as indicators for lightning on Titan". *Proceedings of the Second European Workshop on Exo/Astrobiology, Graz, Australia, 16–19 September*.
- [54] Morente J. A., Molina-Cuberos G. J., Porti J. A., K. Schwingenschuh, B. P. Besser (2003). "A study of the propagation of electromagnetic waves in Titan's atmosphere with the TLM numerical method". *Icarus* **162** (2): 374–384. Bibcode 2003Icar..162..374M. doi:10.1016/S0019-1035(03)00025-3.
- [55] Molina-Cuberos G. J., J. Porti, B. P. Besser, J. A. Morente, J. Margineda, H. I. M. Lichtenegger, A. Salinas, K. Schwingenschuh, H. U. Eichelberger (2004). "Schumann resonances and electromagnetic transparency in the atmosphere of Titan". *Advances in Space Research* **33** (12): 2309–2313. Bibcode 2004AdSpR..33.2309M. doi:10.1016/S0273-1177(03)00465-4.
- [56] Nickolaenko A. P., B. P. Besser, K. Schwingenschuh (2003). "Model computations of Schumann resonance on Titan". *Planet. Space Sci.* **51** (13): 853–862. Bibcode 2003P&SS...51..853N. doi:10.1016/S0032-0633(03)00119-3.
- [57] Bar-Nun A. (1975). "Thunderstorms on Jupiter". *Icarus* **24**: 86–94. Bibcode 1975Icar...24...86B. doi:10.1016/0019-1035(75)90162-1.
- [58] Sentman D. D. (1990). "Electrical conductivity of Jupiter's Shallow interior and the formation of a resonant planetary-ionosphere cavity". *Icarus* **88**: 73–86. Bibcode 1990Icar...88...73S. doi:10.1016/0019-1035(90)90177-B.

## External articles and references

### General references

- Articles on the NASA ADS Database: Full list ([http://adsabs.harvard.edu/cgi-bin/nph-abs\\_connect?db\\_key=AST&db\\_key=PHY&db\\_key=PRE&qform=AST&sim\\_query=YES&ned\\_query=YES&aut\\_logic=OR&obj\\_logic=OR&author=&object=&start\\_mon=&start\\_year=&end\\_mon=&end\\_year=&tlt\\_logic=OR&title="Schumann+Resonances"&txt\\_logic=OR&text=&nr\\_to\\_return=100&start\\_nr=1&jou\\_pick=ALL&ref\\_stems=&data\\_and=ALL&group\\_and=ALL&start\\_entry\\_day=&start\\_entry\\_mon=&start\\_entry\\_year=&end\\_entry\\_day=&end\\_entry\\_mon=&end\\_entry\\_year=&min\\_score=&sort=SCORE&data\\_type=SHORT&aut\\_syn=YES&tlt\\_syn=YES&txt\\_syn=YES&aut\\_wt=1.0&obj\\_wt=1.0&tlt\\_wt=0.3&txt\\_wt=3.0&aut\\_wgt=YES&obj\\_wgt=YES&tlt\\_wgt=YES&txt\\_wgt=YES&tlt\\_sco=YES&txt\\_sco=YES&version=1](http://adsabs.harvard.edu/cgi-bin/nph-abs_connect?db_key=AST&db_key=PHY&db_key=PRE&qform=AST&sim_query=YES&ned_query=YES&aut_logic=OR&obj_logic=OR&author=&object=&start_mon=&start_year=&end_mon=&end_year=&tlt_logic=OR&title=)) | Full text ([http://adsabs.harvard.edu/cgi-bin/nph-abs\\_connect?db\\_key=AST&db\\_key=PHY&db\\_key=PRE&qform=AST&sim\\_query=YES&ned\\_query=YES&aut\\_logic=OR&obj\\_logic=OR&author=&object=&start\\_mon=&start\\_year=&end\\_mon=&end\\_year=&tlt\\_logic=OR&title="Schumann+Resonances"&txt\\_logic=OR&text=&nr\\_to\\_return=100&start\\_nr=1&jou\\_pick=ALL&ref\\_stems=&data\\_and=ALL&group\\_and=ALL&start\\_entry\\_day=&start\\_entry\\_mon=&start\\_entry\\_year=&end\\_entry\\_day=&end\\_entry\\_mon=&end\\_entry\\_year=&min\\_score=&sort=SCORE&data\\_type=SHORT&aut\\_syn=YES&tlt\\_syn=YES&txt\\_syn=YES&aut\\_wt=1.0&obj\\_wt=1.0&tlt\\_wt=0.3&txt\\_wt=3.0&aut\\_wgt=YES&obj\\_wgt=YES&tlt\\_wgt=YES&txt\\_wgt=YES&tlt\\_sco=YES&txt\\_sco=YES&version=1](http://adsabs.harvard.edu/cgi-bin/nph-abs_connect?db_key=AST&db_key=PHY&db_key=PRE&qform=AST&sim_query=YES&ned_query=YES&aut_logic=OR&obj_logic=OR&author=&object=&start_mon=&start_year=&end_mon=&end_year=&tlt_logic=OR&title=))

### Websites

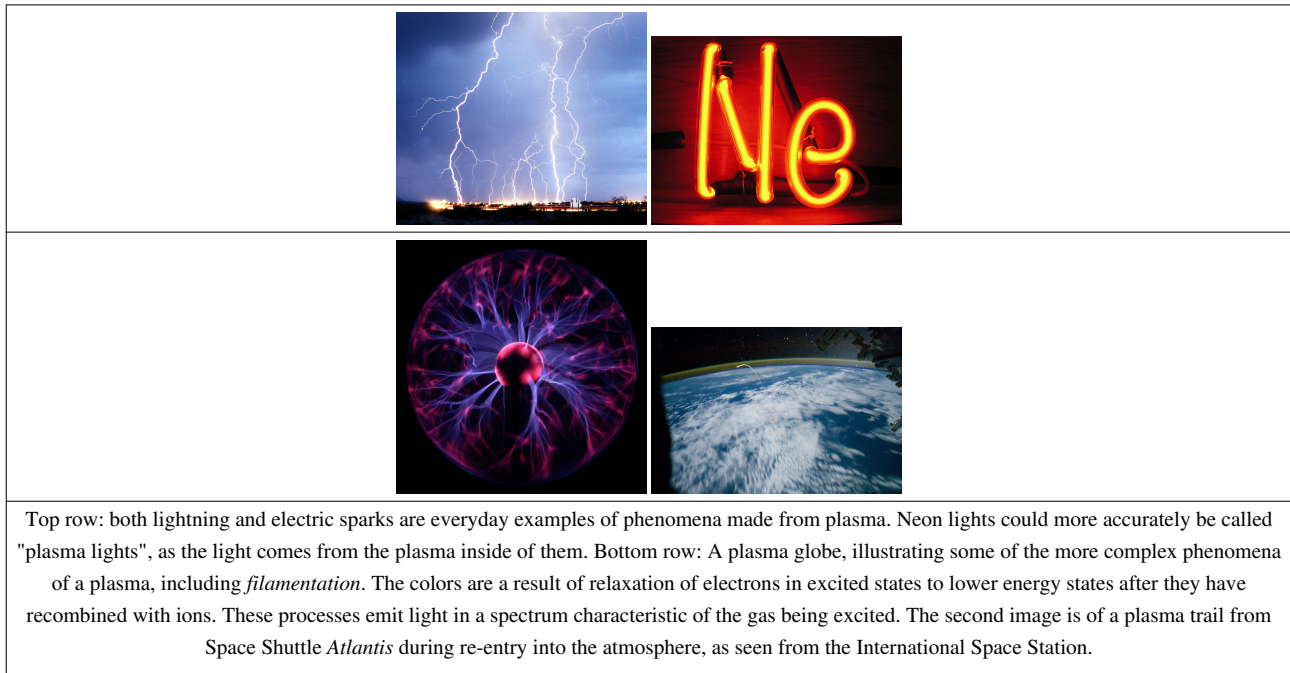
- Magnetic activity and Schumann resonance (<http://quake.geo.berkeley.edu/ncedc/em.intro.html>)
- Well illustrated study from the University of Iowa explaining the construction of a ULF receiver for studying Schumann resonances. Link updated 3 Nov. 2011 ([http://www.iaspacegrant.org/sites/default/files/pdf\\_files/Kruger-SEED.pdf](http://www.iaspacegrant.org/sites/default/files/pdf_files/Kruger-SEED.pdf))
- Global Coherence Initiative (Spectrogram Calendar) (<http://www.glcoherence.org/monitoring-system/live-data.html>) Schumann resonance live data

### Animation

- Schumann resonance animation (<http://svs.gsfc.nasa.gov/vis/a010000/a010800/a010891/>) from NASA Goddard Space Flight Center

# Plasma (physics)

## Plasma



**Plasma** (from Greek *πλάσμα*, "anything formed"<sup>[1]</sup>) is one of the four fundamental states of matter (the others being solid, liquid, and gas). Heating a gas may ionize its molecules or atoms (reducing or increasing the number of electrons in them), thus turning it into a plasma, which contains charged particles: positive ions and negative electrons or ions.<sup>[2]</sup> Ionization can be induced by other means, such as strong electromagnetic field applied with a laser or microwave generator, and is accompanied by the dissociation of molecular bonds, if present.<sup>[3]</sup>

The presence of a non-negligible number of charge carriers makes the plasma electrically conductive so that it responds strongly to electromagnetic fields. Plasma, therefore, has properties quite unlike those of solids, liquids, or gases and is considered a distinct state of matter. Like gas, plasma does not have a definite shape or a definite volume unless enclosed in a container; unlike gas, under the influence of a magnetic field, it may form structures such as filaments, beams and double layers. Some common plasmas are found in stars and neon signs. In the universe, plasma is the most common state of matter for ordinary matter, most of which is in the rarefied intergalactic plasma (particularly intracluster medium) and in stars. Much of the understanding of plasmas has come from the pursuit of controlled nuclear fusion and fusion power, for which plasma physics provides the scientific basis.

## Common plasmas

Plasmas are by far the most common phase of ordinary matter in the universe, both by mass and by volume.<sup>[4]</sup> Our Sun, and all the stars are made of plasma, much of interstellar space is filled with a plasma, albeit a very sparse one, and intergalactic space too. In our solar system, interplanetary space is filled with the plasma of the Solar Wind that extends from the Sun out to the heliopause. Even black holes, which are not directly visible, are fuelled by accreting ionising matter (i.e. plasma),<sup>[5]</sup> and they are associated with astrophysical jets of luminous ejected plasma,<sup>[6]</sup> such as M87's jet that extends 5,000 light-years.<sup>[7]</sup>

Dust and small grains within a plasma will also pick up a net negative charge, so that they in turn may act like a very heavy negative ion component of the plasma (see dusty plasmas).

The current consensus is that about 96% of the total energy density in the universe is not plasma or any other form of ordinary matter, but a combination of cold dark matter and dark energy. In our Solar System, however, the density of ordinary matter is much higher than average and much higher than that of either dark matter or dark energy. The planet Jupiter accounts for most of the *non*-plasma, only about 0.1% of the mass and 10<sup>-15</sup>% of the volume within the orbit of Pluto.

### Common forms of plasma

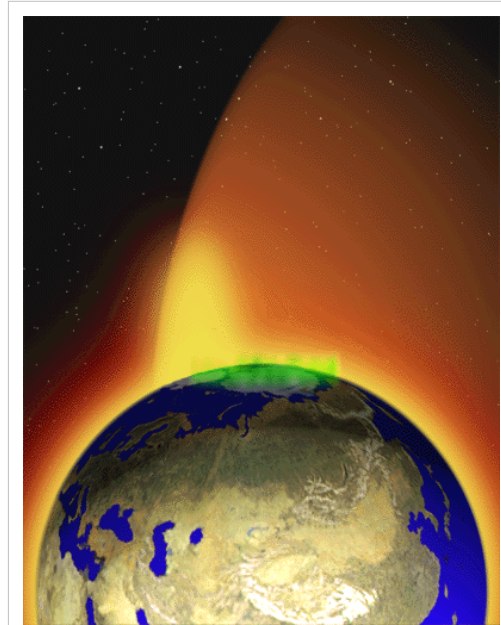
Artificially produced	Terrestrial plasmas	Space and astrophysical plasmas
<ul style="list-style-type: none"> <li>• Those found in plasma displays, including TVs</li> <li>• Inside fluorescent lamps (low energy lighting), neon signs<sup>[8]</sup></li> <li>• Rocket exhaust and ion thrusters</li> <li>• The area in front of a spacecraft's heat shield during re-entry into the atmosphere</li> <li>• Inside a corona discharge ozone generator</li> <li>• Fusion energy research</li> <li>• The electric arc in an arc lamp, an arc welder or plasma torch</li> <li>• Plasma ball (sometimes called a plasma sphere or plasma globe)</li> <li>• Arcs produced by Tesla coils (resonant air core transformer or disruptor coil that produces arcs similar to lightning, but with alternating current rather than static electricity)</li> <li>• Plasmas used in semiconductor device fabrication including reactive-ion etching, sputtering, surface cleaning and plasma-enhanced chemical vapor deposition</li> <li>• Laser-produced plasmas (LPP), found when high power lasers interact with materials.</li> <li>• Inductively coupled plasmas (ICP), formed typically in argon gas for optical emission spectroscopy or mass spectrometry</li> <li>• Magnetically induced plasmas (MIP), typically produced using microwaves as a resonant coupling method</li> <li>• Static electric sparks</li> </ul>	<ul style="list-style-type: none"> <li>• Lightning</li> <li>• St. Elmo's fire</li> <li>• Upper-atmospheric lightning (e.g. Blue jets, Blue starters, Gigantic jets, ELVES)</li> <li>• Sprites</li> <li>• The ionosphere</li> <li>• The plasmasphere</li> <li>• The polar aurorae</li> <li>• Some flames<sup>[9][10]</sup></li> <li>• The polar wind, a plasma fountain</li> </ul>	<ul style="list-style-type: none"> <li>• The Sun and other stars (plasmas heated by nuclear fusion)</li> <li>• The solar wind</li> <li>• The interplanetary medium (space between planets)</li> <li>• The interstellar medium (space between star systems)</li> <li>• The Intergalactic medium (space between galaxies)</li> <li>• The Io-Jupiter flux tube</li> <li>• Accretion discs</li> <li>• Interstellar nebulae</li> <li>• Cometary ion tail</li> </ul>

## Plasma properties and parameters

### Definition of a plasma

Plasma is loosely described as an electrically neutral medium of positive and negative particles (i.e. the overall charge of a plasma is roughly zero). It is important to note that although they are unbound, these particles are not 'free'. When the charges move they generate electrical currents with magnetic fields, and as a result, they are affected by each other's fields. This governs their collective behavior with many degrees of freedom.<sup>[3][12]</sup> A definition can have three criteria:<sup>[13][14]</sup>

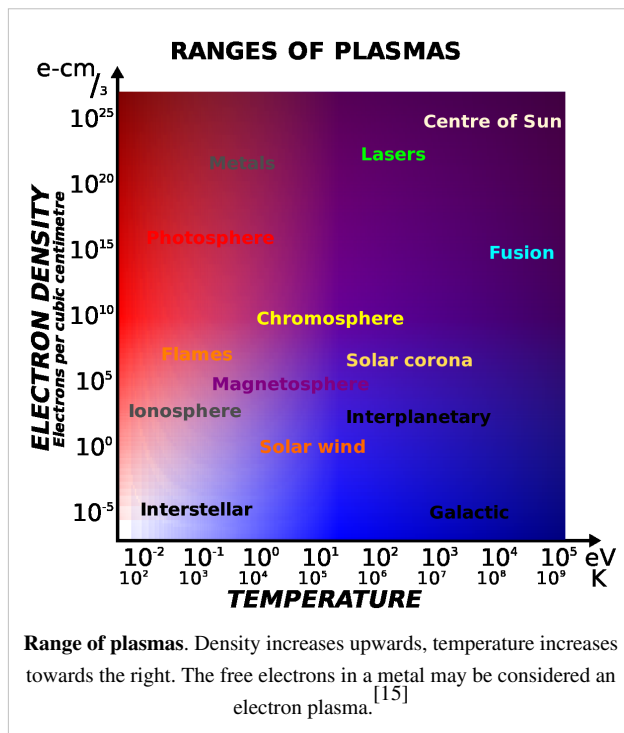
1. **The plasma approximation:** Charged particles must be close enough together that each particle influences many nearby charged particles, rather than just interacting with the closest particle (these collective effects are a distinguishing feature of a plasma). The plasma approximation is valid when the number of charge carriers within the sphere of influence (called the *Debye sphere* whose radius is the Debye screening length) of a particular particle is higher than unity to provide collective behavior of the charged particles. The average number of particles in the Debye sphere is given by the plasma parameter, " $\Lambda$ " (the Greek letter Lambda).
2. **Bulk interactions:** The Debye screening length (defined above) is short compared to the physical size of the plasma. This criterion means that interactions in the bulk of the plasma are more important than those at its edges, where boundary effects may take place. When this criterion is satisfied, the plasma is quasineutral.
3. **Plasma frequency:** The electron plasma frequency (measuring plasma oscillations of the electrons) is large compared to the electron-neutral collision frequency (measuring frequency of collisions between electrons and neutral particles). When this condition is valid, electrostatic interactions dominate over the processes of ordinary gas kinetics.



Artist's rendition of the Earth's plasma fountain, showing oxygen, helium, and hydrogen ions that gush into space from regions near the Earth's poles. The faint yellow area shown above the north pole represents gas lost from Earth into space; the green area is the aurora borealis, where plasma energy pours back into the atmosphere.<sup>[11]</sup>

### Ranges of plasma parameters

Plasma parameters can take on values varying by many orders of magnitude, but the properties of plasmas with apparently disparate parameters may be very similar (see plasma scaling). The following chart considers only conventional atomic plasmas and not exotic phenomena like quark gluon plasmas:



### Typical ranges of plasma parameters: orders of magnitude (OOM)

Characteristic	Terrestrial plasmas	Cosmic plasmas
<b>Size</b> in meters	$10^{-6}$ m (lab plasmas) to $10^2$ m (lightning) (~8 OOM)	$10^{-6}$ m (spacecraft sheath) to $10^{25}$ m (intergalactic nebula) (~31 OOM)
<b>Lifetime</b> in seconds	$10^{-12}$ s (laser-produced plasma) to $10^7$ s (fluorescent lights) (~19 OOM)	$10^1$ s (solar flares) to $10^{17}$ s (intergalactic plasma) (~16 OOM)
<b>Density</b> in particles per cubic meter	$10^7$ m <sup>-3</sup> to $10^{32}$ m <sup>-3</sup> (inertial confinement plasma)	$1$ m <sup>-3</sup> (intergalactic medium) to $10^{30}$ m <sup>-3</sup> (stellar core)
<b>Temperature</b> in kelvins	~0 K (crystalline non-neutral plasma <sup>[16]</sup> ) to $10^8$ K (magnetic fusion plasma)	$10^2$ K (aurora) to $10^7$ K (solar core)
<b>Magnetic fields</b> in teslas	$10^{-4}$ T (lab plasma) to $10^3$ T (pulsed-power plasma)	$10^{-12}$ T (intergalactic medium) to $10^{11}$ T (near neutron stars)

## Degree of ionization

For plasma to exist, ionization is necessary. The term "plasma density" by itself usually refers to the "electron density", that is, the number of free electrons per unit volume. The degree of ionization of a plasma is the proportion of atoms that have lost or gained electrons, and is controlled mostly by the temperature. Even a partially ionized gas in which as little as 1% of the particles are ionized can have the characteristics of a plasma (i.e., response to magnetic fields and high electrical conductivity). The degree of ionization,  $\alpha$  is defined as  $\alpha = n_i / (n_i + n_a)$  where  $n_i$  is the number density of ions and  $n_a$  is the number density of neutral atoms. The *electron density* is related to this by the average charge state  $\langle Z \rangle$  of the ions through  $n_e = \langle Z \rangle n_i$  where  $n_e$  is the number density of electrons.

## Temperatures

Plasma temperature is commonly measured in kelvins or electronvolts and is, informally, a measure of the thermal kinetic energy per particle. Very high temperatures are usually needed to sustain ionization, which is a defining feature of a plasma. The degree of plasma ionization is determined by the "electron temperature" relative to the ionization energy, (and more weakly by the density), in a relationship called the Saha equation. At low temperatures, ions and electrons tend to recombine into bound states—atoms,<sup>[17]</sup> and the plasma will eventually become a gas.

In most cases the electrons are close enough to thermal equilibrium that their temperature is relatively well-defined, even when there is a significant deviation from a Maxwellian energy distribution function, for example, due to UV radiation, energetic particles, or strong electric fields. Because of the large difference in mass, the electrons come to thermodynamic equilibrium amongst themselves much faster than they come into equilibrium with the ions or neutral atoms. For this reason, the "ion temperature" may be very different from (usually lower than) the "electron temperature". This is especially common in weakly ionized technological plasmas, where the ions are often near the ambient temperature.

## Thermal vs. non-thermal plasmas

Based on the relative temperatures of the electrons, ions and neutrals, plasmas are classified as "thermal" or "non-thermal". Thermal plasmas have electrons and the heavy particles at the same temperature, i.e., they are in thermal equilibrium with each other. Non-thermal plasmas on the other hand have the ions and neutrals at a much lower temperature (normally room temperature), whereas electrons are much "hotter".

A plasma is sometimes referred to as being "hot" if it is nearly fully ionized, or "cold" if only a small fraction (for example 1%) of the gas molecules are ionized, but other definitions of the terms "hot plasma" and "cold plasma" are common. Even in a "cold" plasma, the electron temperature is still typically several thousand degrees Celsius. Plasmas utilized in "plasma technology" ("technological plasmas") are usually cold in the sense that only a small fraction of the gas molecules are ionized.

## Potentials

Since plasmas are very good conductors, electric potentials play an important role. The potential as it exists on average in the space between charged particles, independent of the question of how it can be measured, is called the "plasma potential", or the "space potential". If an electrode is inserted into a plasma, its potential will generally lie considerably below the plasma potential due to what is termed a Debye sheath. The good electrical conductivity of plasmas makes their electric fields very small. This results in the important concept of "quasineutrality", which says the density of negative charges is approximately equal to the density of positive charges over large volumes of the plasma ( $n_e = \langle Z \rangle n_i$ ), but on the scale of the Debye length there can be charge imbalance. In the special case that *double layers* are formed, the charge separation can extend some tens of Debye lengths.

The magnitude of the potentials and electric fields must be determined by means other than simply finding the net charge density. A common example is to assume that the electrons satisfy the "Boltzmann relation":

$$n_e \propto e^{e\Phi/k_B T_e}.$$

Differentiating this relation provides a means to calculate the electric field from the density:

$$\vec{E} = (k_B T_e / e) (\nabla n_e / n_e).$$

It is possible to produce a plasma that is not quasineutral.

An electron beam, for example, has only negative charges. The density of a non-neutral plasma must generally be very low, or it must be very small, otherwise it will be dissipated by the repulsive electrostatic force.

In astrophysical plasmas, Debye screening prevents electric fields from directly affecting the plasma over large distances, i.e., greater than the Debye length. However, the existence of charged particles causes the plasma to generate and can be affected by magnetic fields. This can and does cause extremely complex behavior, such as the generation of plasma double layers, an object that separates charge over a few tens of Debye lengths. The dynamics of plasmas interacting with external and self-generated magnetic fields are studied in the academic discipline of magnetohydrodynamics.

## Magnetization

Plasma with a magnetic field strong enough to influence the motion of the charged particles is said to be magnetized. A common quantitative criterion is that a particle on average completes at least one gyration around the magnetic field before making a collision, i.e.,  $\omega_{ce} / \nu_{coll} > 1$ , where  $\omega_{ce}$  is the "electron gyrofrequency" and  $\nu_{coll}$  is the "electron collision rate". It is often the case that the electrons are magnetized while the ions are not. Magnetized plasmas are *anisotropic*, meaning that their properties in the direction parallel to the magnetic field are different from those perpendicular to it. While electric fields in plasmas are usually small due to the high conductivity, the electric field associated with a plasma moving in a magnetic field is given by  $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$  (where  $\mathbf{E}$  is the electric field,  $\mathbf{v}$  is the



Lightning is an example of plasma present at Earth's surface. Typically, lightning discharges 30,000 amperes at up to 100 million volts, and emits light, radio waves, X-rays and even gamma rays.<sup>[18]</sup> Plasma temperatures in lightning can approach ~28,000 kelvin and electron densities may exceed  $10^{24} \text{ m}^{-3}$ .



velocity, and  $\mathbf{B}$  is the magnetic field), and is not affected by Debye shielding.<sup>[19]</sup>

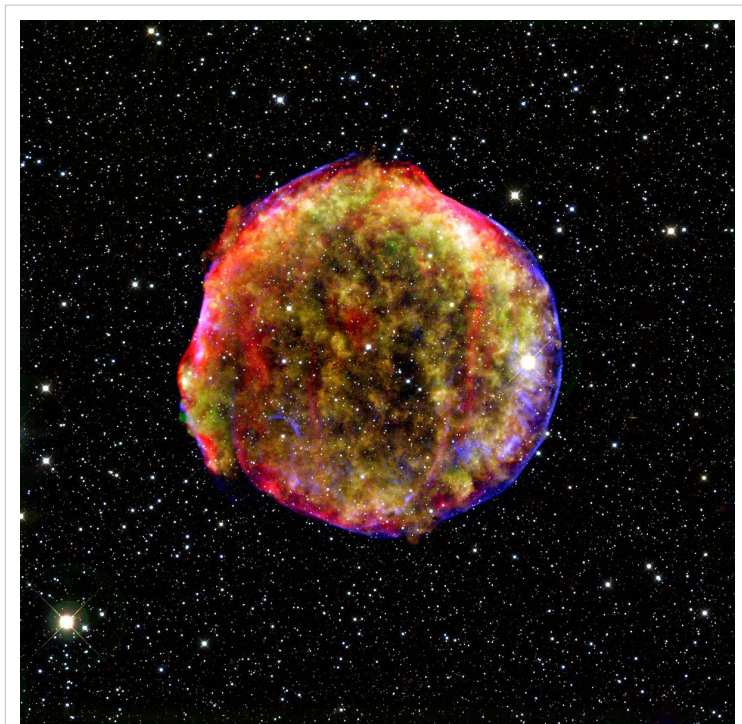
### Comparison of plasma and gas phases

Plasma is often called the *fourth state of matter* after solid, liquids and gases.<sup>[20]</sup> It is distinct from these and other lower-energy states of matter. Although it is closely related to the gas phase in that it also has no definite form or volume, it differs in a number of ways, including the following:

Property	Gas	Plasma
<b>Electrical conductivity</b>	<b>Very low:</b> Air is an excellent insulator until it breaks down into plasma at electric field strengths above 30 kilovolts per centimeter. <sup>[21]</sup>	<b>Usually very high:</b> For many purposes, the conductivity of a plasma may be treated as infinite.
<b>Independently acting species</b>	<b>One:</b> All gas particles behave in a similar way, influenced by gravity and by collisions with one another.	<b>Two or three:</b> Electrons, ions, protons and neutrons can be distinguished by the sign and value of their charge so that they behave independently in many circumstances, with different bulk velocities and temperatures, allowing phenomena such as new types of waves and instabilities.
<b>Velocity distribution</b>	<b>Maxwellian:</b> Collisions usually lead to a Maxwellian velocity distribution of all gas particles, with very few relatively fast particles.	<b>Often non-Maxwellian:</b> Collisional interactions are often weak in hot plasmas and external forcing can drive the plasma far from local equilibrium and lead to a significant population of unusually fast particles.
<b>Interactions</b>	<b>Binary:</b> Two-particle collisions are the rule, three-body collisions extremely rare.	<b>Collective:</b> Waves, or organized motion of plasma, are very important because the particles can interact at long ranges through the electric and magnetic forces.

### Complex plasma phenomena

Although the underlying equations governing plasmas are relatively simple, plasma behavior is extraordinarily varied and subtle: the emergence of unexpected behavior from a simple model is a typical feature of a complex system. Such systems lie in some sense on the boundary between ordered and disordered behavior and cannot typically be described either by simple, smooth, mathematical functions, or by pure randomness. The spontaneous formation of interesting spatial features on a wide range of length scales is one manifestation of plasma complexity. The features are interesting, for example, because they are very sharp, spatially intermittent (the distance between features is much larger than the features themselves), or have a fractal form. Many of these features were first studied in the laboratory, and have subsequently been recognized throughout the universe. Examples of complexity and complex structures in plasmas include:



The remnant of "Tycho's Supernova", a huge ball of expanding plasma. The outer shell shown in blue is X-ray emission by high-speed electrons.



## Filamentation

Striations or string-like structures,<sup>[22]</sup> also known as birkeland currents, are seen in many plasmas, like the plasma ball, the aurora,<sup>[23]</sup> lightning,<sup>[24]</sup> electric arcs, solar flares,<sup>[25]</sup> and supernova remnants.<sup>[26]</sup> They are sometimes associated with larger current densities, and the interaction with the magnetic field can form a magnetic rope structure.<sup>[27]</sup> High power microwave breakdown at atmospheric pressure also leads to the formation of filamentary structures.<sup>[28]</sup> (See also Plasma pinch)

Filamentation also refers to the self-focusing of a high power laser pulse. At high powers, the nonlinear part of the index of refraction becomes important and causes a higher index of refraction in the center of the laser beam, where the laser is brighter than at the edges, causing a feedback that focuses the laser even more. The tighter focused laser has a higher peak brightness (irradiance) that forms a plasma. The plasma has an index of refraction lower than one, and causes a defocusing of the laser beam. The interplay of the focusing index of refraction, and the defocusing plasma makes the formation of a long filament of plasma that can be micrometers to kilometers in length.<sup>[29]</sup> (See also Filament propagation)

## Shocks or double layers

Plasma properties change rapidly (within a few Debye lengths) across a two-dimensional sheet in the presence of a (moving) shock or (stationary) double layer. Double layers involve localized charge separation, which causes a large potential difference across the layer, but does not generate an electric field outside the layer. Double layers separate adjacent plasma regions with different physical characteristics, and are often found in current carrying plasmas. They accelerate both ions and electrons.

## Electric fields and circuits

Quasineutrality of a plasma requires that plasma currents close on themselves in electric circuits. Such circuits follow Kirchhoff's circuit laws and possess a resistance and inductance. These circuits must generally be treated as a strongly coupled system, with the behavior in each plasma region dependent on the entire circuit. It is this strong coupling between system elements, together with nonlinearity, which may lead to complex behavior. Electrical circuits in plasmas store inductive (magnetic) energy, and should the circuit be disrupted, for example, by a plasma instability, the inductive energy will be released as plasma heating and acceleration. This is a common explanation for the heating that takes place in the solar corona. Electric currents, and in particular, magnetic-field-aligned electric currents (which are sometimes generically referred to as "Birkeland currents"), are also observed in the Earth's aurora, and in plasma filaments.

## Cellular structure

Narrow sheets with sharp gradients may separate regions with different properties such as magnetization, density and temperature, resulting in cell-like regions. Examples include the magnetosphere, heliosphere, and heliospheric current sheet. Hannes Alfvén wrote: "From the cosmological point of view, the most important new space research discovery is probably the cellular structure of space. As has been seen in every region of space accessible to in situ measurements, there are a number of 'cell walls', sheets of electric currents, which divide space into compartments with different magnetization, temperature, density, etc."<sup>[30]</sup>

### **Critical ionization velocity**

The critical ionization velocity is the relative velocity between an ionized plasma and a neutral gas, above which a runaway ionization process takes place. The critical ionization process is a quite general mechanism for the conversion of the kinetic energy of a rapidly streaming gas into ionization and plasma thermal energy. Critical phenomena in general are typical of complex systems, and may lead to sharp spatial or temporal features.

### **Ultracold plasma**

Ultracold plasmas are created in a magneto-optical trap (MOT) by trapping and cooling neutral atoms, to temperatures of 1 mK or lower, and then using another laser to ionize the atoms by giving each of the outermost electrons just enough energy to escape the electrical attraction of its parent ion.

One advantage of ultracold plasmas are their well characterized and tunable initial conditions, including their size and electron temperature. By adjusting the wavelength of the ionizing laser, the kinetic energy of the liberated electrons can be tuned as low as 0.1 K, a limit set by the frequency bandwidth of the laser pulse. The ions inherit the millikelvin temperatures of the neutral atoms, but are quickly heated through a process known as disorder induced heating (DIH). This type of non-equilibrium ultracold plasma evolves rapidly, and displays many other interesting phenomena.<sup>[31]</sup>

One of the metastable states of a strongly nonideal plasma is Rydberg matter, which forms upon condensation of excited atoms.

### **Non-neutral plasma**

The strength and range of the electric force and the good conductivity of plasmas usually ensure that the densities of positive and negative charges in any sizeable region are equal ("quasineutrality"). A plasma with a significant excess of charge density, or, in the extreme case, is composed of a single species, is called a non-neutral plasma. In such a plasma, electric fields play a dominant role. Examples are charged particle beams, an electron cloud in a Penning trap and positron plasmas.<sup>[32]</sup>

---

## Dusty plasma and grain plasma

A dusty plasma contains tiny charged particles of dust (typically found in space). The dust particles acquire high charges and interact with each other. A plasma that contains larger particles is called grain plasma. Under laboratory conditions, dusty plasmas are also called *complex plasmas*.<sup>[33]</sup>

## Mathematical descriptions

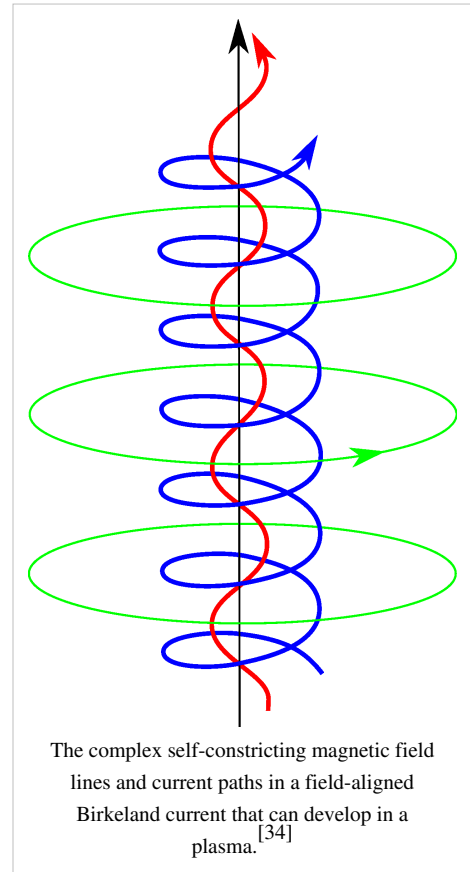
To completely describe the state of a plasma, we would need to write down all the particle locations and velocities and describe the electromagnetic field in the plasma region. However, it is generally not practical or necessary to keep track of all the particles in a plasma. Therefore, plasma physicists commonly use less detailed descriptions, of which there are two main types:

### Fluid model

Fluid models describe plasmas in terms of smoothed quantities, like density and averaged velocity around each position (see Plasma parameters). One simple fluid model, magnetohydrodynamics, treats the plasma as a single fluid governed by a combination of Maxwell's equations and the Navier–Stokes equations. A more general description is the two-fluid plasma picture, where the ions and electrons are described separately. Fluid models are often accurate when collisionality is sufficiently high to keep the plasma velocity distribution close to a Maxwell–Boltzmann distribution. Because fluid models usually describe the plasma in terms of a single flow at a certain temperature at each spatial location, they can neither capture velocity space structures like beams or double layers, nor resolve wave-particle effects.

### Kinetic model

Kinetic models describe the particle velocity distribution function at each point in the plasma and therefore do not need to assume a Maxwell–Boltzmann distribution. A kinetic description is often necessary for collisionless plasmas. There are two common approaches to kinetic description of a plasma. One is based on representing the smoothed distribution function on a grid in velocity and position. The other, known as the particle-in-cell (PIC) technique, includes kinetic information by following the trajectories of a large number of individual particles. Kinetic models are generally more computationally intensive than fluid models. The Vlasov equation may be used to describe the dynamics of a system of charged particles interacting with an electromagnetic field. In magnetized plasmas, a gyrokinetic approach can substantially reduce the computational expense of a fully kinetic simulation.



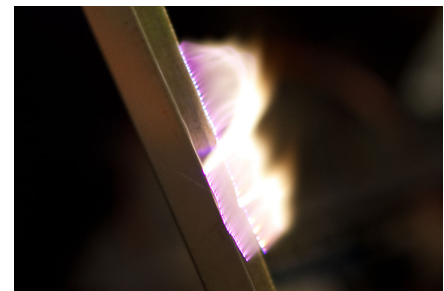
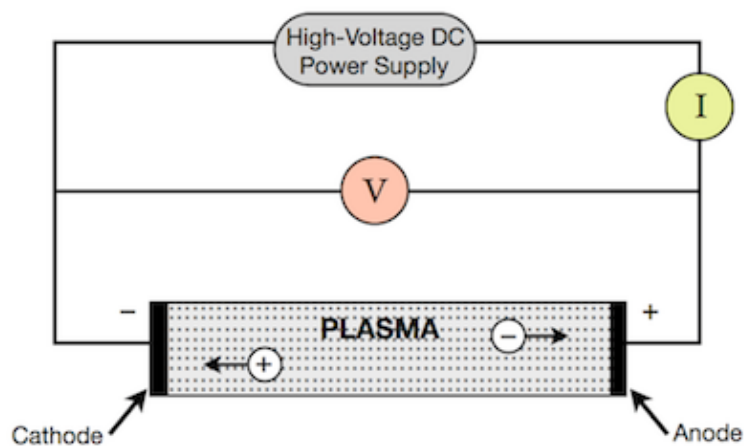
## Artificial plasmas

Most artificial plasmas are generated by the application of electric and/or magnetic fields. Plasma generated in a laboratory setting and for industrial use can be generally categorized by:

- The type of power source used to generate the plasma—DC, RF and microwave
- The pressure they operate at—vacuum pressure ( $< 10$  mTorr or 1 Pa), moderate pressure ( $\sim 1$  Torr or 100 Pa), atmospheric pressure (760 Torr or 100 kPa)
- The degree of ionization within the plasma—fully, partially, or weakly ionized
- The temperature relationships within the plasma—thermal plasma ( $T_e = T_{\text{ion}} = T_{\text{gas}}$ ), non-thermal or "cold" plasma ( $T_e \gg T_{\text{ion}} = T_{\text{gas}}$ )
- The electrode configuration used to generate the plasma
- The magnetization of the particles within the plasma—magnetized (both ion and electrons are trapped in Larmor orbits by the magnetic field), partially magnetized (the electrons but not the ions are trapped by the magnetic field), non-magnetized (the magnetic field is too weak to trap the particles in orbits but may generate Lorentz forces)
- The application

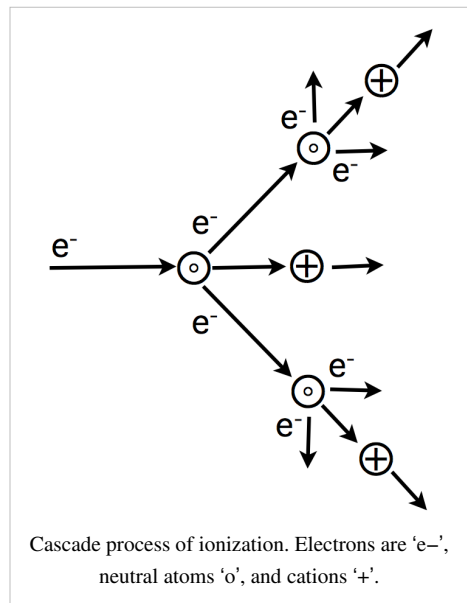
## Generation of artificial plasma

Just like the many uses of plasma, there are several means for its generation, however, one principle is common to all of them: there must be energy input to produce and sustain it.<sup>[35]</sup> For this case, plasma is generated when an electrical current is applied across a dielectric gas or fluid (an electrically non-conducting material) as can be seen in the image below, which shows a discharge tube as a simple example (DC used for simplicity).



Artificial plasma produced in air by a Jacob's Ladder

The potential difference and subsequent electric field pull the bound electrons (negative) toward the anode (positive electrode) while the cathode (negative electrode) pulls the nucleus.<sup>[36]</sup> As the voltage increases, the current stresses the material (by electric polarization) beyond its dielectric limit (termed strength) into a stage of electrical breakdown, marked by an electric spark, where the material transforms from being an insulator into a conductor (as it becomes increasingly ionized). This is a stage of avalanching ionization, where collisions between electrons and neutral gas atoms create more ions and electrons (as can be seen in the figure on the right). The first impact of an electron on an atom results in one ion and two electrons. Therefore, the number of charged particles increases rapidly (in the millions) only “after about 20 successive sets of collisions”,<sup>[1]</sup> mainly due to a small mean free path (average distance travelled between collisions).



### Electric arc

With ample current density and ionization, this forms a luminous electric arc (essentially lightning) between the electrodes.<sup>[37]</sup> Electrical resistance along the continuous electric arc creates heat, which ionizes more gas molecules (where degree of ionization is determined by temperature), and as per the sequence: solid-liquid-gas-plasma, the gas is gradually turned into a thermal plasma.<sup>[38]</sup> A thermal plasma is in thermal equilibrium, which is to say that the temperature is relatively homogeneous throughout the heavy particles (i.e. atoms, molecules and ions) and electrons. This is so because when thermal plasmas are generated, electrical energy is given to electrons, which, due to their great mobility and large numbers, are able to disperse it rapidly and by elastic collision (without energy loss) to the heavy particles.<sup>[39][40]</sup>

### Examples of industrial/commercial plasma

Because of their sizable temperature and density ranges, plasmas find applications in many fields of research, technology and industry. For example, in: industrial and extractive metallurgy,<sup>[39]</sup> surface treatments such as thermal spraying (coating), etching in microelectronics,<sup>[41]</sup> metal cutting<sup>[42]</sup> and welding; as well as in everyday vehicle exhaust cleanup and fluorescent/luminescent lamps,<sup>[35]</sup> while even playing a part in supersonic combustion engines for aerospace engineering.<sup>[43]</sup>

### Low-pressure discharges

- *Glow discharge plasmas*: non-thermal plasmas generated by the application of DC or low frequency RF (<100 kHz) electric field to the gap between two metal electrodes. Probably the most common plasma; this is the type of plasma generated within fluorescent light tubes.<sup>[44]</sup>
- *Capacitively coupled plasma (CCP)*: similar to glow discharge plasmas, but generated with high frequency RF electric fields, typically 13.56 MHz. These differ from glow discharges in that the sheaths are much less intense. These are widely used in the microfabrication and integrated circuit manufacturing industries for plasma etching and plasma enhanced chemical vapor deposition.<sup>[45]</sup>
- *Cascaded Arc Plasma Source*: a device to produce low temperature (~1eV) high density plasmas.
- *Inductively coupled plasma (ICP)*: similar to a CCP and with similar applications but the electrode consists of a coil wrapped around the discharge volume that inductively excites the plasma.
- *Wave heated plasma*: similar to CCP and ICP in that it is typically RF (or microwave), but is heated by both electrostatic and electromagnetic means. Examples are helicon discharge, electron cyclotron resonance (ECR), and ion cyclotron resonance (ICR). These typically require a coaxial magnetic field for wave propagation.

### Atmospheric pressure

- *Arc discharge*: this is a high power thermal discharge of very high temperature (~10,000 K). It can be generated using various power supplies. It is commonly used in metallurgical processes. For example, it is used to melt rocks containing  $\text{Al}_2\text{O}_3$  to produce aluminium.
- *Corona discharge*: this is a non-thermal discharge generated by the application of high voltage to sharp electrode tips. It is commonly used in ozone generators and particle precipitators.
- *Dielectric barrier discharge (DBD)*: this is a non-thermal discharge generated by the application of high voltages across small gaps wherein a non-conducting coating prevents the transition of the plasma discharge into an arc. It is often mislabeled 'Corona' discharge in industry and has similar application to corona discharges. It is also widely used in the web treatment of fabrics.<sup>[46]</sup> The application of the discharge to synthetic fabrics and plastics functionalizes the surface and allows for paints, glues and similar materials to adhere.<sup>[47]</sup>
- *Capacitive discharge*: this is a nonthermal plasma generated by the application of RF power (e.g., 13.56 MHz) to one powered electrode, with a grounded electrode held at a small separation distance on the order of 1 cm. Such discharges are commonly stabilized using a noble gas such as helium or argon.<sup>[48]</sup>

## History

Plasma was first identified in a Crookes tube, and so described by Sir William Crookes in 1879 (he called it "radiant matter").<sup>[49]</sup> The nature of the Crookes tube "cathode ray" matter was subsequently identified by British physicist Sir J.J. Thomson in 1897.<sup>[50]</sup> The term "plasma" was coined by Irving Langmuir in 1928,<sup>[51]</sup> perhaps because the glowing discharge molds itself to the shape of the Crookes tube (Gr. *πλάσμα* – a thing moulded or formed).<sup>[52]</sup> Langmuir described his observations as:

Except near the electrodes, where there are *sheaths* containing very few electrons, the ionized gas contains ions and electrons in about equal numbers so that the resultant space charge is very small. We shall use the name *plasma* to describe this region containing balanced charges of ions and electrons.<sup>[51]</sup>

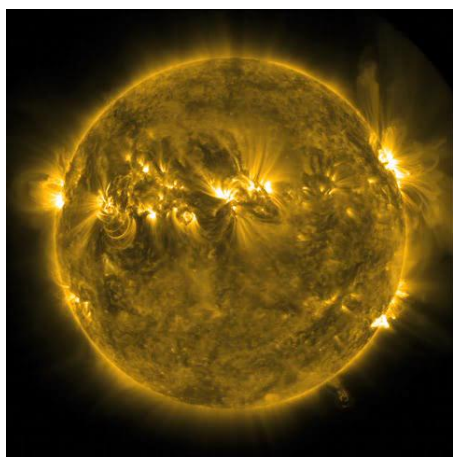
## Fields of active research

This is just a partial list of topics. See list of plasma (physics) articles. A more complete and organized list can be found on the web site Plasma science and technology.<sup>[53]</sup>



Hall effect thruster. The electric field in a plasma double layer is so effective at accelerating ions that electric fields are used in ion drives.

- Plasma theory
  - Plasma equilibria and stability
  - Plasma interactions with waves and beams
  - Guiding center
  - Adiabatic invariant
  - Debye sheath
  - Coulomb collision
- Plasmas in nature
  - The Earth's ionosphere
  - Northern and southern (polar) lights
  - Space plasmas, e.g. Earth's plasmasphere (an inner portion of the magnetosphere dense with plasma)
  - Astrophysical plasma
  - Interplanetary medium
- Industrial plasmas
  - Plasma chemistry
  - Plasma processing
  - Plasma spray
  - Plasma display
  - Plasma sources
  - Dusty plasmas
- Plasma diagnostics
  - Thomson scattering
  - Langmuir probe
  - Spectroscopy
  - Interferometry
  - Ionospheric heating
  - Incoherent scatter radar
- Plasma applications
  - Fusion power
    - Magnetic fusion energy (MFE) — tokamak, stellarator, reversed field pinch, magnetic mirror, dense plasma focus
    - Inertial fusion energy (IFE) (also Inertial confinement fusion — ICF)
    - Plasma-based weaponry
  - Ion implantation
  - Ion thruster
  - MAGPIE (short for *Mega Ampere Generator for Plasma Implosion Experiments*)
  - Plasma ashing
  - Food processing (nonthermal plasma, aka "cold plasma")
  - Plasma arc waste disposal, convert waste into reusable material with plasma.
  - Plasma acceleration
  - Plasma medicine (e. g. Dentistry [54])
  - Plasma window



Solar plasma



Plasma spraying

## Notes

- [1]  $\pi\lambda\sigma\mu\alpha$  (<http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0057:entry=plasma>), Henry George Liddell, Robert Scott, *A Greek-English Lexicon*, on Perseus
- [2] Luo, Q-Z; D'Angelo, N; Merlino, R. L. (1998). *Shock formation in a negative ion plasma* (<http://www.physics.uiowa.edu/~rmerlino/nishocks.pdf>). 5. Department of Physics and Astronomy. . Retrieved 2011-11-20.
- [3] Sturrock, Peter A. (1994). *Plasma Physics: An Introduction to the Theory of Astrophysical, Geophysical & Laboratory Plasmas*. Cambridge University Press. ISBN 0-521-44810-7.
- [4] It is often stated that more than 99% of the material in the visible universe is plasma. See, for example, D. A. Gurnett, A. Bhattacharjee (2005). *Introduction to Plasma Physics: With Space and Laboratory Applications* (<http://books.google.com/?id=VcueZlunrbcC&pg=PA2>). Cambridge, UK: Cambridge University Press. p. 2. ISBN 0-521-36483-3. . and K Scherer, H Fichtner, B Heber (2005). *Space Weather: The Physics Behind a Slogan* (<http://books.google.com/?id=irHgIUtLi0gC&pg=PA138>). Berlin: Springer. p. 138. ISBN 3-540-22907-8. .. Essentially, all of the visible light from space comes from stars, which are plasmas with a temperature such that they radiate strongly at visible wavelengths. Most of the ordinary (or baryonic) matter in the universe, however, is found in the intergalactic medium, which is also a plasma, but much hotter, so that it radiates primarily as X-rays. The current scientific consensus is that about 96% of the total energy density in the universe is not plasma or any other form of ordinary matter, but a combination of cold dark matter and dark energy.
- [5] Péter Mészáros, *The High Energy Universe: Ultra-High Energy Events in Astrophysics and Cosmology*, Publisher Cambridge University Press, 2010, ISBN 0521517001, 780521517003, 222 pages, page 99 ([http://books.google.co.uk/books?id=NXvE\\_zQX5kAC&lpq=PA99&dq="Black hole" plasma accreting&pg=PA99#v=onepage&q="Black hole" plasma accreting&f=false](http://books.google.co.uk/books?id=NXvE_zQX5kAC&lpq=PA99&dq=)).
- [6] Derek J. Raine, Edwin George Thomas, *Black Holes: An Introduction*, Publisher: Imperial College Press, 2010, ISBN 1848163827, 9781848163829, 198 pages, page 160 ([http://books.google.co.uk/books?id=O3puAMw5U3UC&lpq=PA160&dq="black hole" astrophysical jets plasma&pg=PA160#v=onepage&q="black hole" astrophysical jets plasma&f=false](http://books.google.co.uk/books?id=O3puAMw5U3UC&lpq=PA160&dq=))
- [7] Robert Nemiroff (MTU) & Jerry Bonnell (USRA), Astronomy Picture of the Day, 2004 December 2011 (<http://apod.nasa.gov/apod/ap041211.html>), retrieved 9 November 2012
- [8] IPPEX Glossary of Fusion Terms (<http://ippex.pppl.gov/fusion/glossary.html>). Ippex.pppl.gov. Retrieved on 2011-11-19.
- [9] " Plasma and Flames – The Burning Question ([http://www.plasmacoalition.org/plasma\\_writeups/flame.pdf](http://www.plasmacoalition.org/plasma_writeups/flame.pdf))", from the Coalition for Plasma Science, retrieved 8 November 2012
- [10] A. von Engel and J.R. Cozens, "Flame Plasma" in *Advances in electronics and electron physics*, by L. L. Marton, Academic Press, 1976, ISBN 0120145200, 9780120145201, pp. 99 ([http://books.google.co.uk/books?id=0Mndi2cCMuUC&lpq=PA99&ots=fEm9XJ45MD&dq=candle flame plasma&lr&pg=PA99#v=onepage&q=flame plasma&f=false](http://books.google.co.uk/books?id=0Mndi2cCMuUC&lpq=PA99&ots=fEm9XJ45MD&dq=candle+flame+plasma&lr&pg=PA99#v=onepage&q=flame+plasma&f=false))
- [11] Plasma fountain Source (<http://pwg.gsfc.nasa.gov/istp/news/9812/solar1.html>), press release: Solar Wind Squeezes Some of Earth's Atmosphere into Space (<http://pwg.gsfc.nasa.gov/istp/news/9812/solarwind.html>)
- [12] Hazeltine, R.D.; Waelbroeck, F.L. (2004). *The Framework of Plasma Physics*. Westview Press.. ISBN 0-7382-0047-6.
- [13] R. O. Dendy (1990). *Plasma Dynamics* (<http://books.google.com/?id=S1C6-4OBOeYC>). Oxford University Press. ISBN 0-19-852041-7.
- [14] Daniel Hastings, Henry Garrett (2000). *Spacecraft-Environment Interactions*. Cambridge University Press. ISBN 0-521-47128-1.
- [15] Peratt, A. L. (1996). "Advances in Numerical Modeling of Astrophysical and Space Plasmas". *Astrophysics and Space Science* **242** (1–2): 93–163. Bibcode 1996Ap&SS.242...93P. doi:10.1007/BF00645112.
- [16] See The Nonneutral Plasma Group (<http://sdphca.ucsd.edu/>) at the University of California, San Diego
- [17] Nicholson, Dwight R. (1983). *Introduction to Plasma Theory*. John Wiley & Sons. ISBN 0-471-09045-X.
- [18] See Flashes in the Sky: Earth's Gamma-Ray Bursts Triggered by Lightning ([http://www.nasa.gov/vision/universe/solarsystem/rhessi\\_tgf.html](http://www.nasa.gov/vision/universe/solarsystem/rhessi_tgf.html))
- [19] Richard Fitzpatrick, *Introduction to Plasma Physics*, Magnetized plasmas (<http://farside.ph.utexas.edu/teaching/plasma/lectures/node10.html>)
- [20] Yaffa Eliezer, Shalom Eliezer, *The Fourth State of Matter: An Introduction to the Physics of Plasma*, Publisher: Adam Hilger, 1989, ISBN 0852741642, 9780852741641, 226 pages, page 5
- [21] Hong, Alice (2000). "Dielectric Strength of Air" (<http://hypertextbook.com/facts/2000/AliceHong.shtml>). *The Physics Factbook*. .
- [22] Dickel, J. R. (1990). "The Filaments in Supernova Remnants: Sheets, Strings, Ribbons, or?". *Bulletin of the American Astronomical Society* **22**: 832. Bibcode 1990BAAS...22..832D.
- [23] Grydeland, T., *et al.* (2003). "Interferometric observations of filamentary structures associated with plasma instability in the auroral ionosphere". *Geophysical Research Letters* **30** (6): 71. Bibcode 2003GeoRL...30f..71G. doi:10.1029/2002GL016362.
- [24] Moss, Gregory D., *et al.* (2006). "Monte Carlo model for analysis of thermal runaway electrons in streamer tips in transient luminous events and streamer zones of lightning leaders". *Journal of Geophysical Research* **111** (A2): A02307. Bibcode 2006JGRA...11102307M. doi:10.1029/2005JA011350.
- [25] Doherty, Lowell R.; Menzel, Donald H. (1965). "Filamentary Structure in Solar Prominences". *The Astrophysical Journal* **141**: 251. Bibcode 1965ApJ...141..251D. doi:10.1086/148107.
- [26] Hubble views the Crab Nebula M1: The Crab Nebula Filaments ([http://web.archive.org/web/20091005084515/http://seds.lpl.arizona.edu/messier/more/m001\\_hst.html](http://web.archive.org/web/20091005084515/http://seds.lpl.arizona.edu/messier/more/m001_hst.html)). The University of Arizona



- [27] Zhang, Yan-An, *et al.* (2002). "A rope-shaped solar filament and a IIIb flare". *Chinese Astronomy and Astrophysics* **26** (4): 442–450. Bibcode 2002ChA&A..26..442Z. doi:10.1016/S0275-1062(02)00095-4.
- [28] Jean-Pierre Boeuf, Bhaskar Chaudhury, and Guo Qiang Zhu (2010). "Theory and Modeling of Self-Organization and Propagation of Filamentary Plasma Arrays in Microwave Breakdown at Atmospheric Pressure". *Physical Review Letters* **104** (1): 015002. Bibcode 2010PhRvL.104a5002B. doi:10.1103/PhysRevLett.104.015002.
- [29] S. L. Chin (2006). "Some Fundamental Concepts of Femtosecond Laser Filamentation" (<http://icpr.snu.ac.kr/resource/wop.pdf/J01/2006/049/S01/J012006049S010281.pdf>). *Journal of the Korean Physical Society* **49**: 281. .
- [30] Hannes Alfvén (1981). "section VI.13.1. Cellular Structure of Space". *Cosmic Plasma*. Dordrecht. ISBN 90-277-1151-8.
- [31] National Research Council (U.S.). Plasma 2010 Committee (2007). *Plasma science: advancing knowledge in the national interest* (<http://books.google.com/?id=rHo6IbakG2kC&pg=PA190>). National Academies Press. pp. 190–193. ISBN 0-309-10943-4. .
- [32] R. G. Greaves, M. D. Tinkle, and C. M. Surko (1994). "Creation and uses of positron plasmas". *Physics of Plasmas* **1** (5): 1439. Bibcode 1994PhPl....1.1439G. doi:10.1063/1.870693.
- [33] Morfill, G. E. (2009). "Complex plasmas: An interdisciplinary research field". *Review of Modern Physics* **81**: 1353–1404. doi:10.1103/RevModPhys.81.1353.
- [34] See Evolution of the Solar System (<http://history.nasa.gov/SP-345/ch15.htm#250>), 1976)
- [35] Hippler, R., Kersten, H., Schmidt, M., Schoenbach, K.M., ed. (2008). "Plasma Sources". *Low Temperature Plasmas: Fundamentals, Technologies, and Techniques* (2 ed.). Wiley-VCH. ISBN 3-527-40673-5.
- [36] Chen, Francis F. (1984). *Plasma Physics and Controlled Fusion*. Plenum Press. ISBN 0-306-41332-9.
- [37] The material undergoes various 'regimes' or stages (e.g. saturation, breakdown, glow, transition and thermal arc) as the voltage is increased under the voltage-current relationship. The voltage rises to its maximum value in the saturation stage, and thereafter it undergoes fluctuations of the various stages; while the current progressively increases throughout. Leal-Quirós, Edbertho (2004). "Plasma Processing of Municipal Solid Waste". *Brazilian Journal of Physics* **34** (4B): 1587. Bibcode 2004BrJPh...34.1587L.
- [38] Across literature, there appears to be no strict definition on where the boundary is between a gas and plasma. Nevertheless, it is enough to say that at 2000°C the gas molecules become atomized, and ionized at 3000°C and "in this state, [the] gas has a liquid like viscosity at atmospheric pressure and the free electric charges confer relatively high electrical conductivities that can approach those of metals."
- [39] Gomez, E., Rani, D.A., Cheeseman, C.R., Deegan, D., Wise, M., Boccaccini, A.R. (2009). "Thermal plasma technology for the treatment of wastes: A critical review". *Journal of Hazardous Materials* **161** (2–3): 614–626. doi:10.1016/j.jhazmat.2008.04.017. PMID 18499345.
- [40] Note that non-thermal, or non-equilibrium plasmas are not as ionized and have lower energy densities, and thus the temperature is not dispersed evenly among the particles, where some heavy ones remain 'cold'.
- [41] National Research Council (1991). *Plasma Processing of Materials : Scientific Opportunities and Technological Challenges*. National Academies Press. ISBN 0-309-04597-5.
- [42] Nemchinsky, V.A., Severance, W.S. (2006). "What we know and what we do not know about plasma arc cutting". *J. Phys. D: Appl. Phys.* **39** (22): R423–R438. Bibcode 2006JPhD...39R.423N. doi:10.1088/0022-3727/39/22/R01.
- [43] Peretich, M.A., O'Brien, W.F., Schetz, J.A. (2007). *Plasma torch power control for scramjet application* ([http://www.vsgc.odu.edu/src/SRC07/SRC07papers/Mark Peretich \\_ PaperFinal Report.pdf](http://www.vsgc.odu.edu/src/SRC07/SRC07papers/Mark%20Peretich_PaperFinal%20Report.pdf)). Virginia Space Grant Consortium. . Retrieved 12 April 2010.
- [44] Dr. David P. Stern. "The Fluorescent Lamp: A plasma you can use." (<http://www.spof.gsfc.nasa.gov/Education/wfluor.html>). . Retrieved 2010-05-19.
- [45] Sobolewski, M.A.; Langan & Felker, J.G. & B.S. (1997). *Electrical optimization of plasma-enhanced chemical vapor deposition chamber cleaning plasmas* ([http://physics.nist.gov/MajResProj/rfcell/Publications/MAS\\_JVSTB16\\_1.pdf](http://physics.nist.gov/MajResProj/rfcell/Publications/MAS_JVSTB16_1.pdf)). **16**. J. Vac. Sci. Technol. B. pp. 173–182. .
- [46] F. Leroux *et al.* (2006). "Atmospheric air plasma treatments of polyester textile structures". *Journal of Adhesion Science and Technology* **20** (9): 939–957. doi:10.1163/156856106777657788.
- [47] F. Leroux *et al.* (2008). "Polypropylene film chemical and physical modifications by dielectric barrier discharge plasma treatment at atmospheric pressure". *Journal of Colloid and Interface Science* **328** (2): 412. doi:10.1016/j.jcis.2008.09.062. PMID 18930244.
- [48] J. Park *et al.* (2001). "Discharge phenomena of an atmospheric pressure radio-frequency capacitive plasma source". *Journal of Applied Physics* **89** (1): 20. Bibcode 2001JAP....89...20P. doi:10.1063/1.1323753.
- [49] Crookes presented a lecture to the British Association for the Advancement of Science, in Sheffield, on Friday, 22 August 1879 (<http://www.worldcatlibraries.org/wcpa/top3mset/5dcb9349d366f8ec.html>) (<http://www.tfcbooks.com/mall/more/315rm.htm>)
- [50] Announced in his evening lecture to the Royal Institution on Friday, 30th April 1897, and published in *Philosophical Magazine* **44**: 293. 1897. <http://web.lemoyne.edu/~GIUNTA/thomson1897.html>.
- [51] I. Langmuir (1928). "Oscillations in ionized gases". *Proc. Nat. Acad. Sci. U.S.* **14** (8): 628. Bibcode 1928PNAS...14..627L. doi:10.1073/pnas.14.8.627.
- [52] BROWN, Sanborn C. (1978). "Chapter 1: A Short History of Gaseous Electronics". In HIRSH, Merle N. e OSKAM, H. J.. *Gaseous Electronics*. **1**. Academic Press. ISBN 0-12-349701-9.
- [53] Web site for Plasma science and technology (<http://www.plasmas.com/topics.htm>)
- [54] "High-tech dentistry – "St Elmo's frier" – Using a plasma torch to clean your teeth" ([http://www.economist.com/displaystory.cfm?story\\_id=13794903&fsrc=rss](http://www.economist.com/displaystory.cfm?story_id=13794903&fsrc=rss)). The Economist print edition. Jun 17 2009. . Retrieved 2009-09-07.

## References

### External links

- Free plasma physics books and notes (<http://www.freebookcentre.net/Physics/Plasma-Physics-Books.html>)
- Plasmas: the Fourth State of Matter ([http://fusedweb.pppl.gov/CPEP/Chart\\_Pages/5.Plasma4StateMatter.html](http://fusedweb.pppl.gov/CPEP/Chart_Pages/5.Plasma4StateMatter.html))
- Plasma Science and Technology (<http://www.plasmas.org/>)
- Plasma on the Internet (<http://plasma-gate.weizmann.ac.il/directories/plasma-on-the-internet/>) – a list of plasma related links.
- Introduction to Plasma Physics: Graduate course given by Richard Fitzpatrick (<http://farside.ph.utexas.edu/teaching/plasma/lectures/lectures.html>)| M.I.T. Introduction by I.H.Hutchinson (<http://silas.psfc.mit.edu/introplasma/index.html>)
- Plasma Material Interaction (<http://starfire.ne.uiuc.edu/>)
- How to make a glowing ball of plasma in your microwave with a grape (<http://c3po.barnesos.net/homepage/lpl/grapeplasma/>)| More (Video) (<http://stewdio.org/plasma/>)
- How to make plasma in your microwave with only one match (video) (<http://video.google.com/videoplay?docid=6732382807079775486&hl=en>)
- OpenPIC3D – 3D Hybrid Particle-In-Cell simulation of plasma dynamics (<http://comphys.narod.ru>)
- Plasma Formulary Interactive (<http://plasma-gate.weizmann.ac.il/pf/>)

## Space

---

**Space** is the boundless three-dimensional extent in which objects exist and events occur and have relative position and direction.<sup>[1]</sup> Physical space is often conceived in three linear dimensions, although modern physicists usually consider it, with time, to be part of a boundless four-dimensional continuum known as spacetime. In mathematics, "spaces" are examined with different numbers of dimensions and with different underlying structures. The concept of space is considered to be of fundamental importance to an understanding of the physical universe. However, disagreement continues between philosophers over whether it is itself an entity, a relationship between entities, or part of a conceptual framework.

Debates concerning the nature, essence and the mode of existence of space date back to antiquity; namely, to treatises like the *Timaeus* of Plato, or Socrates in his reflections on what the Greeks called *khora* (i.e. "space"), or in the *Physics* of Aristotle (Book IV, Delta) in the definition of *topos* (i.e. place), or even in the later "geometrical conception of place" as "space *qua* extension" in the *Discourse on Place (Qawl fi al-Makan)* of the 11th century Arab polymath Alhazen.<sup>[2]</sup> Many of these classical philosophical questions were discussed in the Renaissance and then reformulated in the 17th century, particularly during the early development of classical mechanics. In Isaac Newton's view, space was absolute—in the sense that it existed permanently and independently of whether there were any matter in the space.<sup>[3]</sup> Other natural philosophers, notably Gottfried Leibniz, thought instead that space was in fact a collection of relations between objects, given by their distance and direction from one another. In the 18th century, the philosopher and theologian George Berkeley attempted to refute the "visibility of spatial depth" in his *Essay Towards a New Theory of Vision*. Later, the metaphysician Immanuel Kant said neither space nor time can be empirically perceived, they are elements of a systematic framework that humans use to structure all experiences. Kant referred to "space" in his *Critique of Pure Reason* as being: a subjective "pure *a priori* form of intuition", hence it is an unavoidable contribution of our human faculties.

In the 19th and 20th centuries mathematicians began to examine non-Euclidean geometries, in which space can be said to be *curved*, rather than *flat*. According to Albert Einstein's theory of general relativity, space around

---

gravitational fields deviates from Euclidean space.<sup>[4]</sup> Experimental tests of general relativity have confirmed that non-Euclidean space provides a better model for the shape of space.

## Philosophy of space

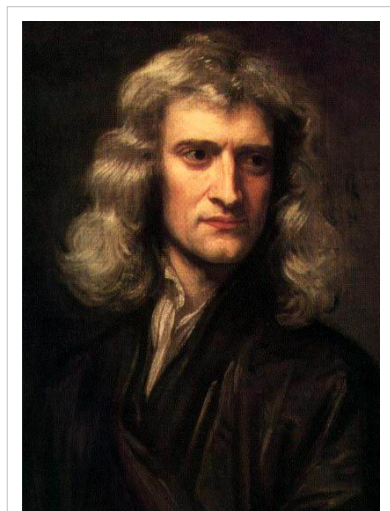
### Leibniz and Newton

In the seventeenth century, the philosophy of space and time emerged as a central issue in epistemology and metaphysics. At its heart, Gottfried Leibniz, the German philosopher-mathematician, and Isaac Newton, the English physicist-mathematician, set out two opposing theories of what space is. Rather than being an entity that independently exists over and above other matter, Leibniz held that space is no more than the collection of spatial relations between objects in the world: "space is that which results from places taken together".<sup>[5]</sup> Unoccupied regions are those that *could* have objects in them, and thus spatial relations with other places. For Leibniz, then, space was an idealised abstraction from the relations between individual entities or their possible locations and therefore could not be continuous but must be discrete.<sup>[6]</sup> Space could be thought of in a similar way to the relations between family members. Although people in the family are related to one another, the relations do not exist independently of the people.<sup>[7]</sup> Leibniz



Gottfried Leibniz

argued that space could not exist independently of objects in the world because that implies a difference between two universes exactly alike except for the location of the material world in each universe. But since there would be no observational way of telling these universes apart then, according to the identity of indiscernibles, there would be no real difference between them. According to the principle of sufficient reason, any theory of space that implied that there could be these two possible universes, must therefore be wrong.<sup>[8]</sup>



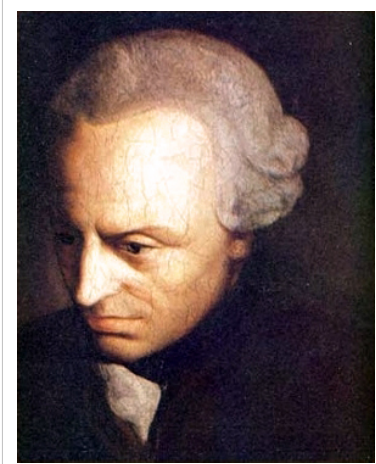
Isaac Newton

Newton took space to be more than relations between material objects and based his position on observation and experimentation. For a relationist there can be no real difference between inertial motion, in which the object travels with constant velocity, and non-inertial motion, in which the velocity changes with time, since all spatial measurements are relative to other objects and their motions. But Newton argued that since non-inertial motion generates forces, it must be absolute.<sup>[9]</sup> He used the example of water in a spinning bucket to demonstrate his argument. Water in a bucket is hung from a rope and set to spin, starts with a flat surface. After a while, as the bucket continues to spin, the surface of the water becomes concave. If the bucket's spinning is stopped then the surface of the water remains concave as it continues to spin. The concave surface is therefore apparently not the result of relative motion between the bucket and the water.<sup>[10]</sup> Instead, Newton argued, it must be a result of non-inertial motion relative to space itself. For several centuries the bucket argument was decisive in showing that space must exist

independently of matter.

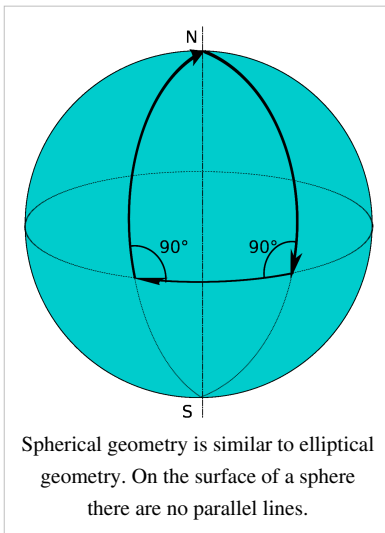
## Kant

In the eighteenth century the German philosopher Immanuel Kant developed a theory of knowledge in which knowledge about space can be both *a priori* and *synthetic*.<sup>[11]</sup> According to Kant, knowledge about space is *synthetic*, in that statements about space are not simply true by virtue of the meaning of the words in the statement. In his work, Kant rejected the view that space must be either a substance or relation. Instead he came to the conclusion that space and time are not discovered by humans to be objective features of the world, but are part of an unavoidable systematic framework for organizing our experiences.<sup>[12]</sup>



Immanuel Kant

## Non-Euclidean geometry



Euclid's *Elements* contained five postulates that form the basis for Euclidean geometry. One of these, the parallel postulate has been the subject of debate among mathematicians for many centuries. It states that on any plane on which there is a straight line  $L_1$  and a point  $P$  not on  $L_1$ , there is only one straight line  $L_2$  on the plane that passes through the point  $P$  and is parallel to the straight line  $L_1$ . Until the 19th century, few doubted the truth of the postulate; instead debate centered over whether it was necessary as an axiom, or whether it was a theory that could be derived from the other axioms.<sup>[13]</sup> Around 1830 though, the Hungarian János Bolyai and the Russian Nikolai Ivanovich Lobachevsky separately published treatises on a type of geometry that does not include the parallel postulate, called hyperbolic geometry. In this geometry, an infinite number of parallel lines pass through the point  $P$ . Consequently the sum of angles in a triangle is less than  $180^\circ$  and the ratio of a circle's circumference to its diameter is greater than pi. In the 1850s,

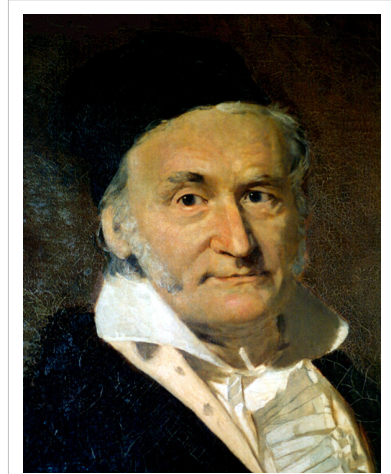
Bernhard Riemann developed an equivalent theory of elliptical geometry, in which no parallel lines pass through  $P$ . In this geometry, triangles have more than  $180^\circ$  and circles have a ratio of circumference-to-diameter that is less than pi.

Type of geometry	Number of parallels	Sum of angles in a triangle	Ratio of circumference to diameter of circle	Measure of curvature
<b>Hyperbolic</b>	Infinite	$< 180^\circ$	$> \pi$	$< 0$
<b>Euclidean</b>	1	$180^\circ$	$\pi$	0
<b>Elliptical</b>	0	$> 180^\circ$	$< \pi$	$> 0$

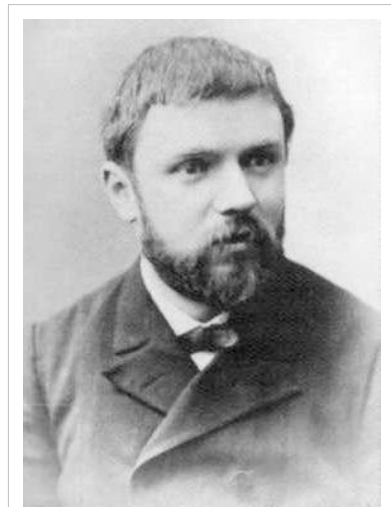
## Gauss and Poincaré

Although there was a prevailing Kantian consensus at the time, once non-Euclidean geometries had been formalised, some began to wonder whether or not physical space is curved. Carl Friedrich Gauss, a German mathematician, was the first to consider an empirical investigation of the geometrical structure of space. He thought of making a test of the sum of the angles of an enormous stellar triangle and there are reports he actually carried out a test, on a small scale, by triangulating mountain tops in Germany.<sup>[14]</sup>

Henri Poincaré, a French mathematician and physicist of the late 19th century introduced an important insight in which he attempted to demonstrate the futility of any attempt to discover which geometry applies to space by experiment.<sup>[15]</sup> He considered the predicament that would face scientists if they were confined to the surface of an imaginary large sphere with particular properties, known as a sphere-world. In this world, the temperature is taken to vary in such a way that all objects expand and contract in similar proportions in different places on the sphere. With a suitable falloff in temperature, if the scientists try to use measuring rods to determine the sum of the angles in a triangle, they can be deceived into thinking that they inhabit a plane, rather than a spherical surface.<sup>[16]</sup> In fact, the scientists cannot in principle determine whether they inhabit a plane or sphere and, Poincaré argued, the same is true for the debate over whether real space is Euclidean or not. For him, which geometry was used to describe space, was a matter of convention.<sup>[17]</sup> Since Euclidean geometry is simpler than non-Euclidean geometry, he assumed the former would always be used to describe the 'true' geometry of the world.<sup>[18]</sup>



Carl Friedrich Gauss

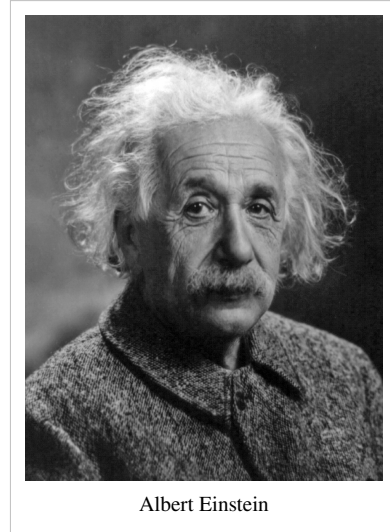


Henri Poincaré

## Einstein

In 1905, Albert Einstein published a paper on a special theory of relativity, in which he proposed that space and time be combined into a single construct known as *spacetime*. In this theory, the speed of light in a vacuum is the same for all observers—which has the result that two events that appear simultaneous to one particular observer will not be simultaneous to another observer if the observers are moving with respect to one another. Moreover, an observer will measure a moving clock to tick more slowly than one that is stationary with respect to them; and objects are measured to be shortened in the direction that they are moving with respect to the observer.

Over the following ten years Einstein worked on a general theory of relativity, which is a theory of how gravity interacts with spacetime. Instead of viewing gravity as a force field acting in spacetime, Einstein suggested that it modifies the geometric structure of spacetime itself.<sup>[19]</sup> According to the general theory, time goes more slowly at places with lower gravitational potentials and rays of light bend in the presence of a gravitational field. Scientists have studied the behaviour of binary pulsars, confirming the predictions of Einstein's theories and non-Euclidean geometry is usually used to describe spacetime.



Albert Einstein

## Mathematics

In modern mathematics spaces are defined as sets with some added structure. They are frequently described as different types of manifolds, which are spaces that locally approximate to Euclidean space, and where the properties are defined largely on local connectedness of points that lie on the manifold. There are however, many diverse mathematical objects that are called spaces. For example, vector spaces such as function spaces may have infinite numbers of independent dimensions and a notion of distance very different to Euclidean space, and topological spaces replace the concept of distance with a more abstract idea of nearness.

## Physics

### Classical mechanics

Space is one of the few fundamental quantities in physics, meaning that it cannot be defined via other quantities because nothing more fundamental is known at the present. On the other hand, it can be related to other fundamental quantities. Thus, similar to other fundamental quantities (like time and mass), space can be explored via measurement and experiment.

### Relativity

Before Einstein's work on relativistic physics, time and space were viewed as independent dimensions. Einstein's discoveries showed that due to relativity of motion our space and time can be mathematically combined into one object — spacetime. It turns out that distances in space or in time separately are not invariant with respect to Lorentz coordinate transformations, but distances in Minkowski space-time along space-time intervals are—which justifies the name.

In addition, time and space dimensions should not be viewed as exactly equivalent in Minkowski space-time. One can freely move in space but not in time. Thus, time and space coordinates are treated differently both in special relativity (where time is sometimes considered an imaginary coordinate) and in general relativity (where different signs are assigned to time and space components of spacetime metric).

Furthermore, in Einstein's general theory of relativity, it is postulated that space-time is geometrically distorted-*curved* -near to gravitationally significant masses.<sup>[20]</sup>

Experiments are ongoing to attempt to directly measure gravitational waves. This is essentially solutions to the equations of general relativity, which describe moving ripples of spacetime. Indirect evidence for this has been found in the motions of the Hulse-Taylor binary system.

## Cosmology

Relativity theory leads to the cosmological question of what shape the universe is, and where space came from. It appears that space was created in the Big Bang, 13.7 billion years ago and has been expanding ever since. The overall shape of space is not known, but space is known to be expanding very rapidly due to the Cosmic Inflation.

## Spatial measurement

The measurement of *physical space* has long been important. Although earlier societies had developed measuring systems, the International System of Units, (SI), is now the most common system of units used in the measuring of space, and is almost universally used.

Currently, the standard space interval, called a standard meter or simply meter, is defined as the distance traveled by light in a vacuum during a time interval of exactly 1/299,792,458 of a second. This definition coupled with present definition of the second is based on the special theory of relativity in which the speed of light plays the role of a fundamental constant of nature.

## Geographical space

Geography is the branch of science concerned with identifying and describing the Earth, utilizing spatial awareness to try to understand why things exist in specific locations. Cartography is the mapping of spaces to allow better navigation, for visualization purposes and to act as a locational device. Geostatistics apply statistical concepts to collected spatial data to create an estimate for unobserved phenomena.

Geographical space is often considered as land, and can have a relation to ownership usage (in which space is seen as property or territory). While some cultures assert the rights of the individual in terms of ownership, other cultures will identify with a communal approach to land ownership, while still other cultures such as Australian Aboriginals, rather than asserting ownership rights to land, invert the relationship and consider that they are in fact owned by the land. Spatial planning is a method of regulating the use of space at land-level, with decisions made at regional, national and international levels. Space can also impact on human and cultural behavior, being an important factor in architecture, where it will impact on the design of buildings and structures, and on farming.

Ownership of space is not restricted to land. Ownership of airspace and of waters is decided internationally. Other forms of ownership have been recently asserted to other spaces—for example to the radio bands of the electromagnetic spectrum or to cyberspace.

Public space is a term used to define areas of land as collectively owned by the community, and managed in their name by delegated bodies; such spaces are open to all, while private property is the land culturally owned by an individual or company, for their own use and pleasure.

Abstract space is a term used in geography to refer to a hypothetical space characterized by complete homogeneity. When modeling activity or behavior, it is a conceptual tool used to limit extraneous variables such as terrain.

## In psychology

Psychologists first began to study the way space is perceived in the middle of the 19th century. Those now concerned with such studies regard it as a distinct branch of psychology. Psychologists analyzing the perception of space are concerned with how recognition of an object's physical appearance or its interactions are perceived, see, for example, visual space.

Other, more specialized topics studied include amodal perception and object permanence. The perception of surroundings is important due to its necessary relevance to survival, especially with regards to hunting and self preservation as well as simply one's idea of personal space.

Several space-related phobias have been identified, including agoraphobia (the fear of open spaces), astrophobia (the fear of celestial space) and claustrophobia (the fear of enclosed spaces).

## References

- [1] Britannica Online Encyclopedia: Space (<http://www.britannica.com/eb/article-9068962/space>)
- [2] Refer to Plato's *Timaeus* in the Loeb Classical Library, Harvard University, and to his reflections on *khora*. See also Aristotle's *Physics*, Book IV, Chapter 5, on the definition of *topos*. Concerning Ibn al-Haytham's 11th century conception of "geometrical place" as "spatial extension", which is akin to Descartes' and Leibniz's 17th century notions of *extensio* and *analysis situs*, and his own mathematical refutation of Aristotle's definition of *topos* in natural philosophy, refer to: Nader El-Bizri, "In Defence of the Sovereignty of Philosophy: al-Baghdadi's Critique of Ibn al-Haytham's Geometrisation of Place", *Arabic Sciences and Philosophy: A Historical Journal* (Cambridge University Press), Vol. 17 (2007), pp. 57-80.
- [3] French and Ebison, *Classical Mechanics*, p. 1
- [4] Carnap, R. *An introduction to the Philosophy of Science*
- [5] Leibniz, Fifth letter to Samuel Clarke
- [6] Vailati, E, Leibniz & Clarke: A Study of Their Correspondence p. 115
- [7] Sklar, L, *Philosophy of Physics*, p. 20
- [8] Sklar, L, *Philosophy of Physics*, p. 21
- [9] Sklar, L, *Philosophy of Physics*, p. 22
- [10] Newton's bucket ([http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Newton\\_bucket.html](http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Newton_bucket.html))
- [11] Carnap, R, *An introduction to the philosophy of science*, p. 177-178
- [12] Lucas, John Randolph. *Space, Time and Causality*. p. 149. ISBN 0-19-875057-9.
- [13] Carnap, R, *An introduction to the philosophy of science*, p. 126
- [14] Carnap, R, *An introduction to the philosophy of science*, p. 134-136
- [15] Jammer, M, *Concepts of Space*, p. 165
- [16] A medium with a variable index of refraction could also be used to bend the path of light and again deceive the scientists if they attempt to use light to map out their geometry
- [17] Carnap, R, *An introduction to the philosophy of science*, p. 148
- [18] Sklar, L, *Philosophy of Physics*, p. 57
- [19] Sklar, L, *Philosophy of Physics*, p. 43
- [20] chapters 8 and 9- John A. Wheeler "A Journey Into Gravity and Spacetime" Scientific American ISBN 0-7167-6034-7



# Chronon

A **chronon** is a proposed quantum of time, that is, a discrete and indivisible "unit" of time as part of a theory that proposes that time is not continuous. While time is a continuous quantity in both standard quantum mechanics and general relativity, many physicists have suggested that a discrete model of time might work, especially when considering the combination of quantum mechanics with general relativity to produce a theory of quantum gravity. The term was introduced in this sense by Robert Lévi.<sup>[1]</sup> Henry Margenau<sup>[2]</sup> suggested that the chronon might be the time for light to travel the classical radius of an electron. A quantum theory in which time is a quantum variable with a discrete spectrum, and which is nevertheless consistent with special relativity, was proposed by Chen Ning Yang.<sup>[3]</sup>

One such model was introduced by Piero Caldirola in 1980. In Caldirola's model, one chronon corresponds to about  $6.97 \times 10^{-24}$  seconds for an electron.<sup>[4]</sup> This is much longer than the Planck time, another proposed unit for the quantization of time, which is only about  $5.39 \times 10^{-44}$  seconds. The Planck time is a universal quantization of time itself, whereas the chronon is a quantization of the evolution in a system along its world line and consequently the value of the chronon, like other quantized observables in quantum mechanics, is a function of the system under consideration, particularly its boundary conditions.<sup>[5]</sup> The value for the chronon,  $\theta_0$ , is calculated from:

$$\theta_0 = \frac{1}{6\pi\epsilon_0} \frac{e^2}{m_0 c^3} \quad [6]$$

From this formula, it can be seen that the nature of the moving particle being considered must be specified since the value of the chronon depends on the particle's charge and mass.

Caldirola claims the chronon has important implications for quantum mechanics, in particular that it allows for a clear answer to the question of whether a free-falling charged particle does or does not emit radiation. This model supposedly avoids the difficulties met by Abraham–Lorentz's and Dirac's approaches to the problem, and provides a natural explication of quantum decoherence.

## Notes

[1] Lévi 1927

[2] Margenau 1950

[3] Yang 1947

[4] Farias & Recami, p.11.

[5] Farias & Recami, p.18.

[6] Farias & Recami, p.11. Caldirola's original paper has a different formula due to not working in standard units.

## References

- Lévi, Robert (1927). "Théorie de l'action universelle et discontinue". *Journal de Physique et le Radium* **8** (4): 182–198. doi:10.1051/jphysrad:0192700804018200.
- Margenau, Henry (1950). *The Nature of Physical Reality*. McGraw-Hill.
- Yang, C N (1947). "On quantized space-time". *Physical Review* **72** (9): 874. Bibcode 1947PhRv...72..874Y. doi:10.1103/PhysRev.72.874.
- Caldirola, P. (1980). "The introduction of the chronon in the electron theory and a charged lepton mass formula". *Lett. Nuovo Cim.* **27** (8): 225–228. doi:10.1007/BF02750348.
- Farias, Ruy A. H.; Recami, Erasmo (1997-06-27). "Introduction of a Quantum of Time ("chronon"), and its Consequences for Quantum Mechanics". arXiv:quant-ph/9706059 [quant-ph].
- Albanese, Claudio; and Lawi, Stephan (2004). "Time Quantization and q-deformations" ([http://eprints.imperial.ac.uk/bitstream/10044/1/48/1/Time Quantization and q.pdf](http://eprints.imperial.ac.uk/bitstream/10044/1/48/1/Time%20Quantization%20and%20q.pdf)). *Journal of Physics A.* **37** (8): 2983–2987. arXiv:hep-th/0308190. Bibcode 2004JPhA...37.2983A. doi:10.1088/0305-4470/37/8/009. Retrieved 2006-07-31.





























# Image Sources, Licenses and Contributors

**File:Scientists montage.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Scientists\\_montage.jpg](http://en.wikipedia.org/w/index.php?title=File:Scientists_montage.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Astrocog, Daniele Pugliesi, JSquish, Mlyard, Muaddib042, 3 anonymous edits

**File:Riccio Melone a Capo Caccia adventurediving.it.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Riccio\\_Melone\\_a\\_Capo\\_Caccia\\_adventurediving.it.jpg](http://en.wikipedia.org/w/index.php?title=File:Riccio_Melone_a_Capo_Caccia_adventurediving.it.jpg) *License:* Creative Commons Attribution 3.0 *Contributors:* Marco Busdraghi

**File:Justus Sustermans - Portrait of Galileo Galilei, 1636.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Justus\\_Sustermans\\_-\\_Portrait\\_of\\_Galileo\\_Galilei\\_1636.jpg](http://en.wikipedia.org/w/index.php?title=File:Justus_Sustermans_-_Portrait_of_Galileo_Galilei_1636.jpg) *License:* Public Domain *Contributors:* Alno, Bukk, Dmitry Rozhkov, G.dallorto, Lupo, Mattes, Meno25, Mysel488, Phrood, Ragesoss, Sercan.ergün, Shakko, Wikiborg4711, Wutsje, 16 anonymous edits

**File:Locke-John-LOC.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Locke-John-LOC.jpg> *License:* Public Domain *Contributors:* Adambro, Dbenbenn, Jengod, Schaengel89, Shizhao, 1 anonymous edits

**File:MenOfScience1807ZobelAndWalker.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:MenOfScience1807ZobelAndWalker.jpg> *License:* Public Domain *Contributors:* by Sir John Gilbert, and Frederick John Skill, and William Walker, and Elizabeth Walker (née Reynolds)

**File:DNA Overview2.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA\\_Overview2.png](http://en.wikipedia.org/w/index.php?title=File:DNA_Overview2.png) *License:* GNU Free Documentation License *Contributors:* Uploader's work on original work by mstroeck

**File:Earth Western Hemisphere.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Earth\\_Western\\_Hemisphere.jpg](http://en.wikipedia.org/w/index.php?title=File:Earth_Western_Hemisphere.jpg) *License:* Public Domain *Contributors:* Col. sheppard, EVula, Juiced lemon, Krenair, Mapmarks, Tom, WikipediaMaster, 3 anonymous edits

**File:Tycho instrument sextant mounting 19.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Tycho\\_instrument\\_sextant\\_mounting\\_19.jpg](http://en.wikipedia.org/w/index.php?title=File:Tycho_instrument_sextant_mounting_19.jpg) *License:* Public Domain *Contributors:* Wikibob

**File:Michelsonmorley-boxplot.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Michelsonmorley-boxplot.svg> *License:* Public Domain *Contributors:* User:Mwtoews, User:Schutz

**File:Hevelius and wife.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Hevelius\\_and\\_wife.jpg](http://en.wikipedia.org/w/index.php?title=File:Hevelius_and_wife.jpg) *License:* Public Domain *Contributors:* Mathiasrex, Matthead, PDH

**File:Meissner effect p1390048.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Meissner\\_effect\\_p1390048.jpg](http://en.wikipedia.org/w/index.php?title=File:Meissner_effect_p1390048.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Mai-Linh Doan

**File:Académie des Sciences 1671.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Académie\\_des\\_Sciences\\_1671.jpg](http://en.wikipedia.org/w/index.php?title=File:Académie_des_Sciences_1671.jpg) *License:* Public Domain *Contributors:* Bohème, JackyR, Karlupart, Man vyi, Mattes, Mdd, Papa November, QWerk, Ragesoss, Shakko, Thib Phil

**File:Estudiante INTEC.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Estudiante\\_INTEC.jpg](http://en.wikipedia.org/w/index.php?title=File:Estudiante_INTEC.jpg) *License:* Creative Commons Attribution-Share Alike *Contributors:* Elenagallegos

**File:NaturalScienceMontage.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:NaturalScienceMontage.png> *License:* Creative Commons Attribution 2.0 *Contributors:* Misc.

**File:Sanzio 01 Plato Aristotle.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Sanzio\\_01\\_Plato\\_Aristotle.jpg](http://en.wikipedia.org/w/index.php?title=File:Sanzio_01_Plato_Aristotle.jpg) *License:* Public Domain *Contributors:* Auntof6, Beria, Bibi Saint-Pol, G.dallorto, Jacobolus, Kentin, Mattes, MonteChristof, Sailko, Tomisti, Wutsje, 5 anonymous edits

**File:Michelson-Morley experiment (en).svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Michelson-Morley\\_experiment\\_\(en\).svg](http://en.wikipedia.org/w/index.php?title=File:Michelson-Morley_experiment_(en).svg) *License:* unknown *Contributors:* User:Bdesham

**File:Moon Dedal crater.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Moon\\_Dedal\\_crater.jpg](http://en.wikipedia.org/w/index.php?title=File:Moon_Dedal_crater.jpg) *License:* Public Domain *Contributors:* Krinkle, Spiritia, Андрей Щербаков, 2 anonymous edits

**File:DNA-fragment-3D-vdW.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:DNA-fragment-3D-vdW.png> *License:* Public Domain *Contributors:* Benjah-bmm27, Bestiasonica

**File:Caffeine.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Caffeine.svg> *License:* Public Domain *Contributors:* Icey, ClockworkSoul

**File:HAtomOrbitals.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:HAtomOrbitals.png> *License:* GNU Free Documentation License *Contributors:* Admrboltz, Benjah-bmm27, Dbc334, Dbenbenn, Ejdzdej, Falcorian, Hongsy, Kborland, MichaelDiederich, Mion, Saperaud, 6 anonymous edits

**File:Ibn al-Haytham.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Ibn\\_al-Haytham.png](http://en.wikipedia.org/w/index.php?title=File:Ibn_al-Haytham.png) *License:* GNU Free Documentation License *Contributors:* derivative work: Wronkiew (talk) Iraqi\_Dinar\_1982\_lbn\_al-Haiitham.jpg: Scan by Jacob L. Bourjaily, Banknote produced by Iraqi government

**Image:Galileo.arp.300pix.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Galileo.arp.300pix.jpg> *License:* Public Domain *Contributors:* ABF, Alefisco, Alno, Bukk, Daniele Pugliesi, David J Wilson, Deadstar, Dirk Hünninger, G.dallorto, Gary King, Herbythyme, Kam Solusar, Liberal Freemason, Michael Bednarek, Phrood, Pérez, Quadell, Ragesoss, Schaengel89, Semnoz, Shakko, Túrelío, Yonatanh, 26 anonymous edits

**File:Johannes Kepler 1610.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Johannes\\_Kepler\\_1610.jpg](http://en.wikipedia.org/w/index.php?title=File:Johannes_Kepler_1610.jpg) *License:* Public Domain *Contributors:* ArtMechanic, Frank C. Müller, Martin H., Mattes, Michael Bednarek, Samuel Grant, Sonphan10, Túrelío, Umherirrender, Xenophon, 4 anonymous edits

**Image:DNA icon (25x25).png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA\\_icon\\_\(25x25\).png](http://en.wikipedia.org/w/index.php?title=File:DNA_icon_(25x25).png) *License:* Public Domain *Contributors:* Bestiasonica, Karelj, Maksim, Rocket000, Wst, 1 anonymous edits

**Image:Jean Louis Théodore Géricault 001.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Jean\\_Louis\\_Théodore\\_Géricault\\_001.jpg](http://en.wikipedia.org/w/index.php?title=File:Jean_Louis_Théodore_Géricault_001.jpg) *License:* Public Domain *Contributors:* AndreasPraefcke, Cgoodwin, Dapal, EDUCA33E, Emijrp, Kersti Nebelsiek, Pitke, Sammyday, Wmpearl, Wst, 1 anonymous edits

**Image:Muybridge race horse animated.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Muybridge\\_race\\_horse\\_animated.gif](http://en.wikipedia.org/w/index.php?title=File:Muybridge_race_horse_animated.gif) *License:* Public Domain *Contributors:* Photos made by Eadweard Muybridge Animation by User Waugsberg

**Image:Perihelion precession.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Perihelion\\_precession.jpg](http://en.wikipedia.org/w/index.php?title=File:Perihelion_precession.jpg) *License:* GNU Free Documentation License *Contributors:* Original uploader was Markus Schmaus at en.wikipedia

**Image:Gravitational lens-full.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Gravitational\\_lens-full.jpg](http://en.wikipedia.org/w/index.php?title=File:Gravitational_lens-full.jpg) *License:* Public Domain *Contributors:* Lukius, Serguei S. Dukachev, 3 anonymous edits

**Image:Aristotle Altemps Inv8575.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Aristotle\\_Altemps\\_Inv8575.jpg](http://en.wikipedia.org/w/index.php?title=File:Aristotle_Altemps_Inv8575.jpg) *License:* Public Domain *Contributors:* User:Jastrow

**File:Roger Bacon.jpeg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Roger\\_Bacon.jpeg](http://en.wikipedia.org/w/index.php?title=File:Roger_Bacon.jpeg) *License:* Public Domain *Contributors:* Dmitry Rozhkov, Elya, Jean-Jacques MILAN, Pieter Kuiper, TommyBee, Tryphon

**File:DNA Structure+Key+Labelled.pn NoBB.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA\\_Structure+Key+Labelled.pn\\_NoBB.png](http://en.wikipedia.org/w/index.php?title=File:DNA_Structure+Key+Labelled.pn_NoBB.png) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Zephyris

**File:ADN animation.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:ADN\\_animation.gif](http://en.wikipedia.org/w/index.php?title=File:ADN_animation.gif) *License:* Public Domain *Contributors:* brian0918

**File:DNA chemical structure.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA\\_chemical\\_structure.svg](http://en.wikipedia.org/w/index.php?title=File:DNA_chemical_structure.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* User:Madprime

**File:DNA orbit animated static thumb.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA\\_orbit\\_animated\\_static\\_thumb.png](http://en.wikipedia.org/w/index.php?title=File:DNA_orbit_animated_static_thumb.png) *License:* GNU Free Documentation License *Contributors:* 84user adapting file originally uploaded by Richard Wheeler (Zephyris) at en.wikipedia

**File:DNA-ligand-by-Abalone.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:DNA-ligand-by-Abalone.png> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* P99am

**File:GC DNA base pair.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:GC\\_DNA\\_base\\_pair.svg](http://en.wikipedia.org/w/index.php?title=File:GC_DNA_base_pair.svg) *License:* Public Domain *Contributors:* Isilanes

**File:AT DNA base pair.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:AT\\_DNA\\_base\\_pair.svg](http://en.wikipedia.org/w/index.php?title=File:AT_DNA_base_pair.svg) *License:* Public Domain *Contributors:* Isilanes

**File:A-DNA, B-DNA and Z-DNA.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:A-DNA,\\_B-DNA\\_and\\_Z-DNA.png](http://en.wikipedia.org/w/index.php?title=File:A-DNA,_B-DNA_and_Z-DNA.png) *License:* GNU Free Documentation License *Contributors:* Original uploader was Richard Wheeler (Zephyris) at en.wikipedia

**File:Parallel telomere quadruple.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Parallel\\_telomere\\_quadruple.png](http://en.wikipedia.org/w/index.php?title=File:Parallel_telomere_quadruple.png) *License:* GNU Free Documentation License *Contributors:* Thomas Spletstoesser

**File:Branch-dna-single.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Branch-dna-single.svg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Otterinfo

**File:Branch-DNA-multiple.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Branch-DNA-multiple.svg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Otterinfo

**File:Cytosin.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Cytosin.svg> *License:* Public Domain *Contributors:* NEUROtiker

**File:5-Methylcytosine.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:5-Methylcytosine.svg> *License:* Public Domain *Contributors:* Yikrazuul (talk)

**File:Thymin.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Thymin.svg> *License:* Public Domain *Contributors:* NEUROtiker



**File:Benzo[a]pyrene DNA adduct 1JJDG.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Benzo\[a\]pyrene\\_DNA\\_adduct\\_1JJDG.png](http://en.wikipedia.org/w/index.php?title=File:Benzo[a]pyrene_DNA_adduct_1JJDG.png) *License:* GNU Free Documentation License *Contributors:* Benjah-bmm27, Bstlee, 1 anonymous edits

**File:T7 RNA polymerase.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:T7\\_RNA\\_polymerase.jpg](http://en.wikipedia.org/w/index.php?title=File:T7_RNA_polymerase.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Splette

**File:DNA replication en.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA\\_replication\\_en.svg](http://en.wikipedia.org/w/index.php?title=File:DNA_replication_en.svg) *License:* Public Domain *Contributors:* LadyofHats Mariana Ruiz

**Image:Nucleosome1.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Nucleosome1.png> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Splette

**File:Lambda repressor ILMB.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Lambda\\_repressor\\_ILMB.png](http://en.wikipedia.org/w/index.php?title=File:Lambda_repressor_ILMB.png) *License:* GNU Free Documentation License *Contributors:* Original uploader was Zephyris at en.wikipedia

**File:EcoRV IRVA.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:EcoRV\\_IRVA.png](http://en.wikipedia.org/w/index.php?title=File:EcoRV_IRVA.png) *License:* GNU Free Documentation License *Contributors:* Original uploader was Zephyris at en.wikipedia

**File:Holliday Junction.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Holliday\\_Junction.svg](http://en.wikipedia.org/w/index.php?title=File:Holliday_Junction.svg) *License:* Public Domain *Contributors:* derivative work: Mouagip (talk) Holliday\_Junction.png: de:User:Chb

**File:Holliday junction coloured.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Holliday\\_junction\\_coloured.png](http://en.wikipedia.org/w/index.php?title=File:Holliday_junction_coloured.png) *License:* GNU Free Documentation License *Contributors:* Original uploader was Zephyris at en.wikipedia

**File:Chromosomal Recombination.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Chromosomal\\_Recombination.svg](http://en.wikipedia.org/w/index.php?title=File:Chromosomal_Recombination.svg) *License:* Creative Commons Attribution 2.5 *Contributors:* David Eccles (Gringer)

**File:DNA nanostructures.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DNA\\_nanostructures.png](http://en.wikipedia.org/w/index.php?title=File:DNA_nanostructures.png) *License:* Creative Commons Attribution 2.5 *Contributors:* (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

**File:Maclyn McCarty with Francis Crick and James D Watson - 10.1371 journal.pbio.0030341.g001-O.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Maclyn\\_McCarty\\_with\\_Francis\\_Crick\\_and\\_James\\_D\\_Watson\\_-\\_10.1371\\_journal.pbio.0030341.g001-O.jpg](http://en.wikipedia.org/w/index.php?title=File:Maclyn_McCarty_with_Francis_Crick_and_James_D_Watson_-_10.1371_journal.pbio.0030341.g001-O.jpg) *License:* Creative Commons Attribution 2.5 *Contributors:* Marjorie McCarty

**File:Acer Aspire 8920 Gemstone by Georgy.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Acer\\_Aspire\\_8920\\_Gemstone\\_by\\_Georgy.JPG](http://en.wikipedia.org/w/index.php?title=File:Acer_Aspire_8920_Gemstone_by_Georgy.JPG) *License:* Trademarked *Contributors:* Georgy90

**File:Columbia Supercomputer - NASA Advanced Supercomputing Facility.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Columbia\\_Supercomputer\\_-\\_NASA\\_Advanced\\_Supercomputing\\_Facility.jpg](http://en.wikipedia.org/w/index.php?title=File:Columbia_Supercomputer_-_NASA_Advanced_Supercomputing_Facility.jpg) *License:* Public Domain *Contributors:* Trower, NASA

**File:Interect Superbrain.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Interect\\_Superbrain.jpg](http://en.wikipedia.org/w/index.php?title=File:Interect_Superbrain.jpg) *License:* Creative Commons Attribution-Sharealike 2.0 *Contributors:* Brighterorange, 1 anonymous edits

**File:2010-01-26-technikkrempele-by-RalfR-05.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:2010-01-26-technikkrempele-by-RalfR-05.jpg> *License:* GNU Free Documentation License *Contributors:* Ralf Roletschek (talk) - Fahrradtechnik auf fahrradmonteur.de

**File:Thinking Machines Connection Machine CM-5 Frostburg 2.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Thinking\\_Machines\\_Connection\\_Machine\\_CM-5\\_Frostburg\\_2.jpg](http://en.wikipedia.org/w/index.php?title=File:Thinking_Machines_Connection_Machine_CM-5_Frostburg_2.jpg) *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Mark Pellegrini

**File:G5 supplying Wikipedia via Gigabit at the Lange Nacht der Wissenschaften 2006 in Dresden.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:G5\\_supplying\\_Wikipedia\\_via\\_Gigabit\\_at\\_the\\_Lange\\_Nacht\\_der\\_Wissenschaften\\_2006\\_in\\_Dresden.JPG](http://en.wikipedia.org/w/index.php?title=File:G5_supplying_Wikipedia_via_Gigabit_at_the_Lange_Nacht_der_Wissenschaften_2006_in_Dresden.JPG) *License:* Creative Commons Attribution 2.5 *Contributors:* Conrad Nutschan

**File:DM IBM S360.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:DM\\_IBM\\_S360.jpg](http://en.wikipedia.org/w/index.php?title=File:DM_IBM_S360.jpg) *License:* Creative Commons Attribution 2.5 *Contributors:* Ben Franske

**File:Acorn BBC Master Series Microcomputer.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Acorn\\_BBC\\_Master\\_Series\\_Microcomputer.jpg](http://en.wikipedia.org/w/index.php?title=File:Acorn_BBC_Master_Series_Microcomputer.jpg) *License:* Creative Commons Attribution-Sharealike 2.0 *Contributors:* MarkusHagenlocher, Mono, Ubcule

**File:Dell PowerEdge Servers.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Dell\\_PowerEdge\\_Servers.jpg](http://en.wikipedia.org/w/index.php?title=File:Dell_PowerEdge_Servers.jpg) *License:* Public Domain *Contributors:* Dsv

**File:Jacquard loom.full.view.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Jacquard\\_loom.full.view.jpg](http://en.wikipedia.org/w/index.php?title=File:Jacquard_loom.full.view.jpg) *License:* Public Domain *Contributors:* User Ghw on en.wikipedia

**File:Jacquard Joseph Marie woven silk.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Jacquard\\_Joseph\\_Marie\\_woven\\_silk.jpg](http://en.wikipedia.org/w/index.php?title=File:Jacquard_Joseph_Marie_woven_silk.jpg) *License:* Public Domain *Contributors:* Astrochemist, Ecummenic, Ezdr, Kilom691, Mdd, WikipediaMaster

**File:Z3 Deutsches Museum.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Z3\\_Deutsches\\_Museum.JPG](http://en.wikipedia.org/w/index.php?title=File:Z3_Deutsches_Museum.JPG) *License:* GNU Free Documentation License *Contributors:* Original uploader was Venusianer at de.wikipedia (Original text : Venusianer 14:13, 3. Jan. 2007 (CET))

**File:Classic shot of the ENIAC.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Classic\\_shot\\_of\\_the\\_ENIAC.jpg](http://en.wikipedia.org/w/index.php?title=File:Classic_shot_of_the_ENIAC.jpg) *License:* Public Domain *Contributors:* Unidentified U.S. Army photographer

**File:EDSAC (10).jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:EDSAC\\_\(10\).jpg](http://en.wikipedia.org/w/index.php?title=File:EDSAC_(10).jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* Copyright Computer Laboratory, University of Cambridge. Reproduced by permission.

**File:80486dx2-large.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:80486dx2-large.jpg> *License:* GNU Free Documentation License *Contributors:* A23cd-s, Adambro, Admrboltz, Artnerisa, CarolSpears, Denniss, Greudin, Julia W, Kozuch, Martin Kozák, Mattbuck, Rjd0060, Rocket000, 11 anonymous edits

**File:SSEM Manchester museum.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:SSEM\\_Manchester\\_museum.jpg](http://en.wikipedia.org/w/index.php?title=File:SSEM_Manchester_museum.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Parrot of Doom

**File:H96566k.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:H96566k.jpg> *License:* Public Domain *Contributors:* Courtesy of the Naval Surface Warfare Center, Dahlgren, VA., 1988.

**File:FortranCardPROJ039.agr.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:FortranCardPROJ039.agr.jpg> *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Arnold Reinhold

**File:Mips32 addi.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Mips32\\_addi.svg](http://en.wikipedia.org/w/index.php?title=File:Mips32_addi.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* en:User:Booyabazooka

**File:Magnetic core.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Magnetic\\_core.jpg](http://en.wikipedia.org/w/index.php?title=File:Magnetic_core.jpg) *License:* Creative Commons Attribution 2.5 *Contributors:* Apalsola, Feyenatic london, Gribozavr, Uberpenguin

**File:HDDspin.JPG** *Source:* <http://en.wikipedia.org/w/index.php?title=File:HDDspin.JPG> *License:* Creative Commons Attribution-Sharealike 2.0 *Contributors:* Alpha six from Germany

**File:Cray 2 Arts et Metiers dsc03940.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Cray\\_2\\_Arts\\_et\\_Metiers\\_dsc03940.jpg](http://en.wikipedia.org/w/index.php?title=File:Cray_2_Arts_et_Metiers_dsc03940.jpg) *License:* Creative Commons Attribution-Sharealike 2.0 *Contributors:* User:David.Monnaix

**File:Internet map 1024.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Internet\\_map\\_1024.jpg](http://en.wikipedia.org/w/index.php?title=File:Internet_map_1024.jpg) *License:* Creative Commons Attribution 2.5 *Contributors:* Barrett Lyon The Opte Project

**File:Car radio antenna extended portrait.jpeg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Car\\_radio\\_antenna\\_extended\\_portrait.jpeg](http://en.wikipedia.org/w/index.php?title=File:Car_radio_antenna_extended_portrait.jpeg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Zuzu

**File:Felder um Dipol.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Felder\\_um\\_Dipol.jpg](http://en.wikipedia.org/w/index.php?title=File:Felder_um_Dipol.jpg) *License:* GNU Free Documentation License *Contributors:* Averse

**File:Canberra Deep Dish Communications Complex - GPN-2000-000502.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Canberra\\_Deep\\_Dish\\_Communications\\_Complex\\_-\\_GPN-2000-000502.jpg](http://en.wikipedia.org/w/index.php?title=File:Canberra_Deep_Dish_Communications_Complex_-_GPN-2000-000502.jpg) *License:* Public Domain *Contributors:* NASA

**File:Antenna.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Antenna.jpg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Yonatan Horan

**File:Rabbit-ears dipole antenna with UHF loop 20090204.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Rabbit-ears\\_dipole\\_antenna\\_with\\_UHF\\_loop\\_20090204.jpg](http://en.wikipedia.org/w/index.php?title=File:Rabbit-ears_dipole_antenna_with_UHF_loop_20090204.jpg) *License:* Creative Commons Attribution 2.5 *Contributors:* Mark Wagner (User:Carnildo)

**File:6 sector site in CDMA.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:6\\_sector\\_site\\_in\\_CDMA.jpg](http://en.wikipedia.org/w/index.php?title=File:6_sector_site_in_CDMA.jpg) *License:* Public Domain *Contributors:* hardikvasa

**File:Westnet wireless cellsite.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Westnet\\_wireless\\_cellsite.JPG](http://en.wikipedia.org/w/index.php?title=File:Westnet_wireless_cellsite.JPG) *License:* Public Domain *Contributors:* Mobileman2009

**File:TV antenna.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:TV\\_antenna.JPG](http://en.wikipedia.org/w/index.php?title=File:TV_antenna.JPG) *License:* Creative Commons Attribution 3.0 *Contributors:* Krish Dulal

**File:Bundesarchiv Bild 183-29802-0001, MTS Strehla, Bezirk Dresden, Ukw-Sprechfunk.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Bundesarchiv\\_Bild\\_183-29802-0001\\_MTS\\_Strehla\\_Bezirk\\_Dresden\\_Ukw-Sprechfunk.jpg](http://en.wikipedia.org/w/index.php?title=File:Bundesarchiv_Bild_183-29802-0001_MTS_Strehla_Bezirk_Dresden_Ukw-Sprechfunk.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 Germany *Contributors:* Chetvorno, Cqdx, El., Fljmm, Mattes, 1 anonymous edits



**File:Supernovastyle Tx Muehlacker.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Supernovastyle\\_Tx\\_Muehlacker.JPG](http://en.wikipedia.org/w/index.php?title=File:Supernovastyle_Tx_Muehlacker.JPG) *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Hans-Peter Scholz, Birkenfeld (Enzkreis), Germany

**File:Folded dipole.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Folded\\_dipole.jpg](http://en.wikipedia.org/w/index.php?title=File:Folded_dipole.jpg) *License:* Creative Commons Attribution 3.0 *Contributors:* Bidgee

**File:Antenna visalia california.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Antenna\\_visalia\\_california.jpg](http://en.wikipedia.org/w/index.php?title=File:Antenna_visalia_california.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Henryk KotowskiKotowski

**File:2008-07-28 Mast radiator.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:2008-07-28\\_Mast\\_radiator.jpg](http://en.wikipedia.org/w/index.php?title=File:2008-07-28_Mast_radiator.jpg) *License:* Creative Commons Attribution-Share Alike *Contributors:* Ildar Sagdejev (Specious)

**File:The Atacama Large Millimeter submillimeter Array (ALMA) by night under the Magellanic Clouds.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:The\\_Atacama\\_Large\\_Millimeter\\_submillimeter\\_Array\\_\(ALMA\)\\_by\\_night\\_under\\_the\\_Magellanic\\_Clouds.jpg](http://en.wikipedia.org/w/index.php?title=File:The_Atacama_Large_Millimeter_submillimeter_Array_(ALMA)_by_night_under_the_Magellanic_Clouds.jpg) *License:* unknown *Contributors:* Jmencison

**File:Half – Wave Dipole.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Half\\_-\\_Wave\\_Dipole.jpg](http://en.wikipedia.org/w/index.php?title=File:Half_-_Wave_Dipole.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Schwarzbeck Mess-Elektronik

**Image:Sidelobes.en.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Sidelobes.en.svg> *License:* Creative Commons Attribution-Share Alike *Contributors:* Timothy Truckle

**File:TVAerial.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:TVAerial.jpg> *License:* Public Domain *Contributors:* Daniel Christensen (talk). Original uploader was Daniel Christensen at en.wikipedia

**File:Old rabbit ears.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Old\\_rabbit\\_ears.jpg](http://en.wikipedia.org/w/index.php?title=File:Old_rabbit_ears.jpg) *License:* Creative Commons Attribution 3.0 *Contributors:* Daniel Christensen (talk)

**File:A6-1EN.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:A6-1EN.jpg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* LP, Santosga

**File:A6-2.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:A6-2.jpg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Anonymous Dissident, Inductiveload, LP, Santosga, 1 anonymous edits

**File:A6-4.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:A6-4.jpg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Elgewen, Jacopo Werther, LP, Santosga, 1 anonymous edits

**File:Zij-en.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Zij-en.png> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Alejo2083, LP, 1 anonymous edits

**Image:Montreal-tower-top.thumb2.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Montreal-tower-top.thumb2.jpg> *License:* Public Domain *Contributors:* Original uploader was Aarchiba at en.wikipedia

**Image:Antenna d44ac.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Antenna\\_d44ac.jpg](http://en.wikipedia.org/w/index.php?title=File:Antenna_d44ac.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Adamantios, ChrisRuvolo, Croquant, HenkvD, Kotowski, Liftarn, Siebrand, Stepa, Waldir

**Image:Television Antenna.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Television\\_Antenna.jpg](http://en.wikipedia.org/w/index.php?title=File:Television_Antenna.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Andreas -horn- Hornig, Conti, Daniel Christensen, Enochlau, Jerome Charles Potts, Lotsa

**Image:Space diversity.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Space\\_diversity.gif](http://en.wikipedia.org/w/index.php?title=File:Space_diversity.gif) *License:* GNU Free Documentation License *Contributors:* David Jordan, Jim.henderson

**Image:LadderlineW3NP.JPG** *Source:* <http://en.wikipedia.org/w/index.php?title=File:LadderlineW3NP.JPG> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* LuckyLouie

**Image:Low cost DCF77 receiver.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Low\\_cost\\_DCF77\\_receiver.jpg](http://en.wikipedia.org/w/index.php?title=File:Low_cost_DCF77_receiver.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Jahoe

**File:136 to 174 MHz base station antennas.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:136\\_to\\_174\\_MHz\\_base\\_station\\_antennas.jpg](http://en.wikipedia.org/w/index.php?title=File:136_to_174_MHz_base_station_antennas.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* David Jordan on English Wikipedia

**Image:VHF UHF LP-antenna.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:VHF\\_UHF\\_LP-antenna.JPG](http://en.wikipedia.org/w/index.php?title=File:VHF_UHF_LP-antenna.JPG) *License:* Attribution *Contributors:* Copyright ©2008 K. Krallis, SV1XV

**Image:Delano VOA.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Delano\\_VOA.jpg](http://en.wikipedia.org/w/index.php?title=File:Delano_VOA.jpg) *License:* Public Domain *Contributors:* Akitoki7 (talk)

**Image:OldTVAntenna.JPG** *Source:* <http://en.wikipedia.org/w/index.php?title=File:OldTVAntenna.JPG> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Sreerambh

**Image:T2FD Antenna.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:T2FD\\_Antenna.png](http://en.wikipedia.org/w/index.php?title=File:T2FD_Antenna.png) *License:* Public Domain *Contributors:* Spahnog

**File:Aerial antenna.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Aerial\\_antenna.JPG](http://en.wikipedia.org/w/index.php?title=File:Aerial_antenna.JPG) *License:* Creative Commons Attribution 3.0 *Contributors:* Daniel Christensen (talk)

**File:Philco am loop.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Philco\\_am\\_loop.jpg](http://en.wikipedia.org/w/index.php?title=File:Philco_am_loop.jpg) *License:* Creative Commons Attribution 3.0 *Contributors:* Daniel Christensen (talk)

**Image:Doncastertower.JPG** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Doncastertower.JPG> *License:* GNU Free Documentation License *Contributors:* Andrew Amos / Skyscraper297 at en.wikipedia

**File:Palmerston-water-tank.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Palmerston-water-tank.jpg> *License:* Creative Commons Attribution 3.0 *Contributors:* Bidgee

**Image:base station mexico-city.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Base\\_station\\_mexico-city.JPG](http://en.wikipedia.org/w/index.php?title=File:Base_station_mexico-city.JPG) *License:* Public Domain *Contributors:* Pptudela

**Image:PalmCellTower.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:PalmCellTower.jpg> *License:* Public Domain *Contributors:* Gary Minnaert (Minnaert)

**Image:Trunked 5ch central control.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Trunked\\_5ch\\_central\\_control.svg](http://en.wikipedia.org/w/index.php?title=File:Trunked_5ch_central_control.svg) *License:* GNU Free Documentation License *Contributors:* Traced by User:Stammered from an original by David Jordan

**Image:Base station antenna network.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Base\\_station\\_antenna\\_network.svg](http://en.wikipedia.org/w/index.php?title=File:Base_station_antenna_network.svg) *License:* Public Domain *Contributors:* Original uploader was David Jordan at en.wikipedia. Later version(s) were uploaded by MashrurR at en.wikipedia.

**Image:Chimp Brain in a jar.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Chimp\\_Brain\\_in\\_a\\_jar.jpg](http://en.wikipedia.org/w/index.php?title=File:Chimp_Brain_in_a_jar.jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* Gaetan Lee . Tilt corrected by Kaldari.

**File:Section through olfactory bulb 16 days old rat brain.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Section\\_through\\_olfactory\\_bulb\\_16\\_days\\_old\\_rat\\_brain.jpg](http://en.wikipedia.org/w/index.php?title=File:Section_through_olfactory_bulb_16_days_old_rat_brain.jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* Prepared for Commons by User:OldakQuill from a CC-by-2.0 figure in a journal article by Rebecca Elsaesser and Jacques Paysan (see source).

**File:Chemical synapse schema cropped.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Chemical\\_synapse\\_schema\\_cropped.jpg](http://en.wikipedia.org/w/index.php?title=File:Chemical_synapse_schema_cropped.jpg) *License:* Public Domain *Contributors:* user:Looie496 created file, US National Institutes of Health, National Institute on Aging created original

**File:PLoSBiol4.e126.Fig6fNeuron.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:PLoSBiol4.e126.Fig6fNeuron.jpg> *License:* Creative Commons Attribution 2.5 *Contributors:* Wei-Chung Allen Lee, Hayden Huang, Guoping Feng, Joshua R. Sanes, Emery N. Brown, Peter T. So, Elly Nedivi

**Image:Bilaterian-plan.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Bilaterian-plan.svg> *License:* Public Domain *Contributors:* Looie496

**Image:Drosophila melanogaster - side (aka).jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Drosophila\\_melanogaster\\_-\\_side\\_\(aka\).jpg](http://en.wikipedia.org/w/index.php?title=File:Drosophila_melanogaster_-_side_(aka).jpg) *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* André Karwath aka Aka

**Image:Shark brain.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Shark\\_brain.png](http://en.wikipedia.org/w/index.php?title=File:Shark_brain.png) *License:* Public Domain *Contributors:* William Skaggs

**Image:EmbryonicBrain.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:EmbryonicBrain.svg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Nrets, vector conversion by Surachit

**Image:Vertebrate-brain-regions small.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Vertebrate-brain-regions\\_small.png](http://en.wikipedia.org/w/index.php?title=File:Vertebrate-brain-regions_small.png) *License:* Public Domain *Contributors:* Vertebrate-brain-regions.png: Looie496 derivative work: Looie496 (talk)

**File:Spike-waves.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Spike-waves.png> *License:* Creative Commons Attribution-Sharealike 2.0 *Contributors:* Der Lange, Lipothymia, Magnus Manske, NEUROtiker, Str4nd, 3 anonymous edits

**File:Model of Cerebellar Perceptron.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Model\\_of\\_Cerebellar\\_Perceptron.jpg](http://en.wikipedia.org/w/index.php?title=File:Model_of_Cerebellar_Perceptron.jpg) *License:* Public Domain *Contributors:* Albus, J.S., NIST

**File:Hearing mechanics cropped.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Hearing\\_mechanics\\_cropped.jpg](http://en.wikipedia.org/w/index.php?title=File:Hearing_mechanics_cropped.jpg) *License:* Public Domain *Contributors:* Zina Deretsky, National Science Foundation

**File:LocationOfHypothalamus.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:LocationOfHypothalamus.jpg> *License:* Public Domain *Contributors:* EhJj, OldakQuill, Was a bee, Poman Беккеп

**File:Basal ganglia.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Basal\\_ganglia.svg](http://en.wikipedia.org/w/index.php?title=File:Basal_ganglia.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Mikael Häggström and Andrew Gillies

**File:6 week embryo brain.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:6\\_week\\_embryo\\_brain.jpg](http://en.wikipedia.org/w/index.php?title=File:6_week_embryo_brain.jpg) *License:* Public Domain *Contributors:* Kurzon

**File:EEG mit 32 Electroden.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:EEG\\_mit\\_32\\_Electroden.jpg](http://en.wikipedia.org/w/index.php?title=File:EEG_mit_32_Electroden.jpg) *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Aschoeke

**File:BCI.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:BCI.jpg> *License:* Creative Commons Attribution 2.5 *Contributors:* see source

**File:Descartes-reflex.JPG** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Descartes-reflex.JPG> *License:* Public Domain *Contributors:* René Descartes

**File:1543,AndreasVesalius'Fabrica,BaseOfTheBrain.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:1543,AndreasVesalius'Fabrica,BaseOfTheBrain.jpg> *License:* Public Domain *Contributors:* User Ancheta Wis on en.wikipedia

**File:PurkinjeCell.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:PurkinjeCell.jpg> *License:* Public Domain *Contributors:* Chrislb, Feezil, Hystrix, Interpretix, Jacklee, SriMesh, 1 anonymous edits

**Image:Neuron Hand-tuned.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Neuron\\_Hand-tuned.svg](http://en.wikipedia.org/w/index.php?title=File:Neuron_Hand-tuned.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Original uploader was Quasar Jarosz at en.wikipedia

**Image:Complete neuron cell diagram en.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Complete\\_neuron\\_cell\\_diagram\\_en.svg](http://en.wikipedia.org/w/index.php?title=File:Complete_neuron_cell_diagram_en.svg) *License:* Public Domain *Contributors:* LadyofHats

**Image:Gyrus Dentatus 40x.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Gyrus\\_Dentatus\\_40x.jpg](http://en.wikipedia.org/w/index.php?title=File:Gyrus_Dentatus_40x.jpg) *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* MethoxyRoxy

**Image:GFPneuron.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:GFPneuron.png> *License:* Creative Commons Attribution 2.5 *Contributors:* Original uploader was Nrets at en.wikipedia

**Image:smi32neuron.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Smi32neuron.jpg> *License:* Creative Commons Attribution 3.0 *Contributors:* UC Regents Davis campus

**File:Neurons uni bi multi pseudouni.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Neurons\\_uni\\_bi\\_multi\\_pseudouni.svg](http://en.wikipedia.org/w/index.php?title=File:Neurons_uni_bi_multi_pseudouni.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Pseudounipolar\_bipolar\_neurons.svg; Jooj8 derivative work: Jonathan Haas (talk)

**Image:Chemical synapse schema cropped.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Chemical\\_synapse\\_schema\\_cropped.jpg](http://en.wikipedia.org/w/index.php?title=File:Chemical_synapse_schema_cropped.jpg) *License:* Public Domain *Contributors:* user:Looie496 created file, US National Institutes of Health, National Institute on Aging created original

**Image:Golgi Hippocampus.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Golgi\\_Hippocampus.jpg](http://en.wikipedia.org/w/index.php?title=File:Golgi_Hippocampus.jpg) *License:* Public Domain *Contributors:* Ephraim33, Marc Gabriel Schmid, Methoxyroxy, Rory096, Was a bee, 2 anonymous edits

**Image:Purkinje cell by Cajal.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Purkinje\\_cell\\_by\\_Cajal.png](http://en.wikipedia.org/w/index.php?title=File:Purkinje_cell_by_Cajal.png) *License:* Public Domain *Contributors:* Original uploader was PNG crusade bot at en.wikipedia

**Image:Mandelbrot-similar-x1.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Mandelbrot-similar-x1.jpg> *License:* GNU General Public License *Contributors:* D-Kuru, Maksim

**Image:Mandelbrot-similar-x6.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Mandelbrot-similar-x6.jpg> *License:* GNU General Public License *Contributors:* D-Kuru, Maksim

**Image:Mandelbrot-similar-x100.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Mandelbrot-similar-x100.jpg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* D-Kuru, Maksim, 2 anonymous edits

**Image:Mandelbrot-similar-x2000.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Mandelbrot-similar-x2000.jpg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* D-Kuru, Maksim, 2 anonymous edits

**Image:Von Koch curve.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Von\\_Koch\\_curve.gif](http://en.wikipedia.org/w/index.php?title=File:Von_Koch_curve.gif) *License:* Public Domain *Contributors:* António Miguel de Campos

**Image:Julia set (indigo).png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Julia\\_set\\_\(indigo\).png](http://en.wikipedia.org/w/index.php?title=File:Julia_set_(indigo).png) *License:* Public Domain *Contributors:* Adam majewski, Conti, D-Kuru, Nk, Saperaud

**Image:Karperien Strange Attractor 200.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Karperien\\_Strange\\_Attractor\\_200.gif](http://en.wikipedia.org/w/index.php?title=File:Karperien_Strange_Attractor_200.gif) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Akarpe

**File:Uniform Triangle Mass Center grade 5 fractal.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Uniform\\_Triangle\\_Mass\\_Center\\_grade\\_5\\_fractal.gif](http://en.wikipedia.org/w/index.php?title=File:Uniform_Triangle_Mass_Center_grade_5_fractal.gif) *License:* Creative Commons Attribution-Share Alike *Contributors:* Ignaciokriche

**File:KarperienFractalBranch.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:KarperienFractalBranch.jpg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Akarpe

**File:Finite subdivision of a radial link.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Finite\\_subdivision\\_of\\_a\\_radial\\_link.png](http://en.wikipedia.org/w/index.php?title=File:Finite_subdivision_of_a_radial_link.png) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Brirush

**File:Apophysis-100303-104.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Apophysis-100303-104.jpg> *License:* Public Domain *Contributors:* Gut Monk (talk)

**Image:Frost patterns 2.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Frost\\_patterns\\_2.jpg](http://en.wikipedia.org/w/index.php?title=File:Frost_patterns_2.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Schnobby

**Image:Glue1 800x600.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Glue1\\_800x600.jpg](http://en.wikipedia.org/w/index.php?title=File:Glue1_800x600.jpg) *License:* GNU Free Documentation License *Contributors:* Darapti, Ies, Paddy, Saperaud, Ttrung, WayneRay, 1 anonymous edits

**Image:Square1.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Square1.jpg> *License:* Attribution *Contributors:* Bert Hickman

**Image:Fractal Broccoli.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Fractal\\_Broccoli.jpg](http://en.wikipedia.org/w/index.php?title=File:Fractal_Broccoli.jpg) *License:* Public Domain *Contributors:* Jon Sullivan

**File:Animated fractal mountain.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Animated\\_fractal\\_mountain.gif](http://en.wikipedia.org/w/index.php?title=File:Animated_fractal_mountain.gif) *License:* Public Domain *Contributors:* António Miguel de Campos - en>User:Tó campos

**File:FibonacciBlocks.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:FibonacciBlocks.svg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Borb, 1 anonymous edits

**File:Fibonacci spiral 34.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Fibonacci\\_spiral\\_34.svg](http://en.wikipedia.org/w/index.php?title=File:Fibonacci_spiral_34.svg) *License:* Public Domain *Contributors:* User:Dicklyon

**Image:Liber abbaçi magliab f124r.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Liber\\_abbaçi\\_magliab\\_f124r.jpg](http://en.wikipedia.org/w/index.php?title=File:Liber_abbaçi_magliab_f124r.jpg) *License:* Public Domain *Contributors:* G.dallorto, Karl.Kirst, Mdd, Otfried Lieberknecht, Sailko, Warburg, 1 anonymous edits

**File:PascalTriangleFibonacci.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:PascalTriangleFibonacci.svg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* RDBury

**File:FibonacciChamomile.PNG** *Source:* <http://en.wikipedia.org/w/index.php?title=File:FibonacciChamomile.PNG> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* User:Alvesgaspar; derivative work: RDBury (talk)

**File:SunflowerModel.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:SunflowerModel.svg> *License:* GNU Free Documentation License *Contributors:* Doron

**File:Golden ratio line.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Golden\\_ratio\\_line.svg](http://en.wikipedia.org/w/index.php?title=File:Golden_ratio_line.svg) *License:* Public Domain *Contributors:* Traced by User:Stannered

**File:SimilarGoldenRectangles.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:SimilarGoldenRectangles.svg> *License:* Creative Commons Zero *Contributors:* Ahecht (Original); Pbroks13 (Derivative work)

**File:Phi uc lc.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Phi\\_uc\\_lc.svg](http://en.wikipedia.org/w/index.php?title=File:Phi_uc_lc.svg) *License:* Public Domain *Contributors:* Dcoetzee, F l a n k e r

**File:Michael Maestlin.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Michael\\_Maestlin.jpg](http://en.wikipedia.org/w/index.php?title=File:Michael_Maestlin.jpg) *License:* Public Domain *Contributors:* original uploader

**File:Acropolis of Athens 01361.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Acropolis\\_of\\_Athens\\_01361.JPG](http://en.wikipedia.org/w/index.php?title=File:Acropolis_of_Athens_01361.JPG) *License:* Public domain *Contributors:* Fred J, Itsmine, Jkelly, LX, Madmedea, 1 anonymous edits

**File:Pentagram and human body (Agrippa).jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Pentagram\\_and\\_human\\_body\\_\(Agrippa\).jpg](http://en.wikipedia.org/w/index.php?title=File:Pentagram_and_human_body_(Agrippa).jpg) *License:* Public Domain *Contributors:* AnonMoos, Mattes, Nyo, Tgr, Waldir, 朝彦, 1 anonymous edits

**File:Medieval manuscript framework.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Medieval\\_manuscript\\_framework.svg](http://en.wikipedia.org/w/index.php?title=File:Medieval_manuscript_framework.svg) *License:* Public Domain *Contributors:* jossi

**File:Aeonium tabuliforme.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Aeonium\\_tabuliforme.jpg](http://en.wikipedia.org/w/index.php?title=File:Aeonium_tabuliforme.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Maxronnersjo

**File:Whirling squares.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Whirling\\_squares.svg](http://en.wikipedia.org/w/index.php?title=File:Whirling_squares.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Dicklyon

**File:Golden mean.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Golden\\_mean.png](http://en.wikipedia.org/w/index.php?title=File:Golden_mean.png) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Adam majewski

**File:FakeRealLogSpiral.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:FakeRealLogSpiral.svg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* FakeRealLogSpiral.png; Pau derivative work: silverhammermba

**File:Goldener Schnitt Konstr beliebt.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Goldener\\_Schnitt\\_Konstr\\_beliebt.svg](http://en.wikipedia.org/w/index.php?title=File:Goldener_Schnitt_Konstr_beliebt.svg) *License:* GNU Free Documentation License *Contributors:* Daniel Seibert Original uploader was Daniel Seibert at de.wikipedia

**File:Golden triangle (math).svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Golden\\_triangle\\_\(math\).svg](http://en.wikipedia.org/w/index.php?title=File:Golden_triangle_(math).svg) *License:* Public Domain *Contributors:* Jheald

**File:Odom.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Odom.svg> *License:* Creative Commons Attribution 3.0 *Contributors:* Kmhknh

**File: Pentagonam-phi.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Pentagram-phi.svg> *License:* Public Domain *Contributors:* Original uploader was Jamiemichelle at en.wikipedia

**File:Ptolemy Pentagon.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Ptolemy\\_Pentagon.svg](http://en.wikipedia.org/w/index.php?title=File:Ptolemy_Pentagon.svg) *License:* Public domain *Contributors:* AnonMoos, Athaenara, JMCC1, Stotr

**File:GoldenRhombus.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:GoldenRhombus.svg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Pbroks13

**File:Rhombictriacontahedron.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Rhombictriacontahedron.svg> *License:* GNU Free Documentation License *Contributors:* User:DTR

**File:Mathematical Pyramid.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Mathematical\\_Pyramid.svg](http://en.wikipedia.org/w/index.php?title=File:Mathematical_Pyramid.svg) *License:* Public Domain *Contributors:* Dicklyon

**File:10 Quantum Mechanics Masters.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:10\\_Quantum\\_Mechanics\\_Masters.jpg](http://en.wikipedia.org/w/index.php?title=File:10_Quantum_Mechanics_Masters.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* derivative work: Patrick Edwin Moran (talk) Einstein1921\_by\_F\_Schmutzer\_2.jpg: Ferdinand Schmutzer (1870-1928) Niels\_Bohr\_Date\_Unverified\_LOC.jpg: unknown Broglie\_Big.jpg: unknown Max\_Born.jpg: unknown Dirac\_4.jpg: unknown Werner\_Heisenberg\_at\_1927\_Solvay\_Conference.JPG: Photograph Institut International de Physique Solvay, Brussels, Belgium Wolfgang\_Pauli\_young.jpg: unknown Richard\_Feynman\_ID\_badge.png: unknown

**File:Hot metalwork.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Hot\\_metalwork.jpg](http://en.wikipedia.org/w/index.php?title=File:Hot_metalwork.jpg) *License:* unknown *Contributors:* Contributor, Fir0002, Jahobr, Wst, 1 anonymous edits

**File:RWP-comparison.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:RWP-comparison.svg> *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* sfu

**File:Einstein.Painting.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Einstein.Painting.jpg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* GeorgeLouis

**File:Photoelectric effect.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Photoelectric\\_effect.svg](http://en.wikipedia.org/w/index.php?title=File:Photoelectric_effect.svg) *License:* GNU Free Documentation License *Contributors:* Wolfmankurd

**File:Emission spectrum-H.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Emission\\_spectrum-H.svg](http://en.wikipedia.org/w/index.php?title=File:Emission_spectrum-H.svg) *License:* Creative Commons Zero *Contributors:* Merikanto, Adrignola

**File:Bohr atom model English.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Bohr\\_atom\\_model\\_English.svg](http://en.wikipedia.org/w/index.php?title=File:Bohr_atom_model_English.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Brighterorange

**File:Young+Fringses.gif** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Young+Fringses.gif> *License:* GNU Free Documentation License *Contributors:* Young.gif: Mpfiz derivative work: Patrick Edwin Moran (talk)

**File:Single & double slit experiment.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Single\\_&\\_double\\_slit\\_experiment.jpg](http://en.wikipedia.org/w/index.php?title=File:Single_&_double_slit_experiment.jpg) *License:* GNU Free Documentation License *Contributors:* Patrick Edwin Moran

**File:Erwin Schroedinger.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Erwin\\_Schroedinger.jpg](http://en.wikipedia.org/w/index.php?title=File:Erwin_Schroedinger.jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:GeorgeLouis

**File:Heisenberg 10.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Heisenberg\\_10.jpg](http://en.wikipedia.org/w/index.php?title=File:Heisenberg_10.jpg) *License:* Public Domain *Contributors:* JdH, Palamède, Quiris

**File:neon orbitals.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Neon\\_orbitals.JPG](http://en.wikipedia.org/w/index.php?title=File:Neon_orbitals.JPG) *License:* Public Domain *Contributors:* Benjah-bmm27, Mortadelo2005, 1 anonymous edits

**File:Dirac 3.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Dirac\\_3.jpg](http://en.wikipedia.org/w/index.php?title=File:Dirac_3.jpg) *License:* anonymous-EU *Contributors:* Cambridge University, Cavendish Laboratory

**File:Superposition.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Superposition.svg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Patrick Edwin Moran

**File:Paul.Dirac.monument.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Paul.Dirac.monument.jpg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* GeorgeLouis, Ww2censor

**Image:Propagation of a de broglie plane wave.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Propagation\\_of\\_a\\_de\\_broglie\\_plane\\_wave.svg](http://en.wikipedia.org/w/index.php?title=File:Propagation_of_a_de_broglie_plane_wave.svg) *License:* Public Domain *Contributors:* User:Maschen

**Image:Propagation of a de broglie wavepacket.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Propagation\\_of\\_a\\_de\\_broglie\\_wavepacket.svg](http://en.wikipedia.org/w/index.php?title=File:Propagation_of_a_de_broglie_wavepacket.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Maschen

**Image:Position and momentum of a Gaussian initial state for a QHO, balanced.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Position\\_and\\_momentum\\_of\\_a\\_Gaussian\\_initial\\_state\\_for\\_a\\_QHO\\_balanced.gif](http://en.wikipedia.org/w/index.php?title=File:Position_and_momentum_of_a_Gaussian_initial_state_for_a_QHO_balanced.gif) *License:* Creative Commons Zero *Contributors:* Teply

**Image:Position and momentum of a Gaussian initial state for a QHO, narrow.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Position\\_and\\_momentum\\_of\\_a\\_Gaussian\\_initial\\_state\\_for\\_a\\_QHO\\_narrow.gif](http://en.wikipedia.org/w/index.php?title=File:Position_and_momentum_of_a_Gaussian_initial_state_for_a_QHO_narrow.gif) *License:* Creative Commons Zero *Contributors:* Teply

**Image:Position and momentum of a Gaussian initial state for a QHO, wide.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Position\\_and\\_momentum\\_of\\_a\\_Gaussian\\_initial\\_state\\_for\\_a\\_QHO\\_wide.gif](http://en.wikipedia.org/w/index.php?title=File:Position_and_momentum_of_a_Gaussian_initial_state_for_a_QHO_wide.gif) *License:* Creative Commons Zero *Contributors:* Teply

**File:Heisenbergbohr.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Heisenbergbohr.jpg> *License:* Public Domain *Contributors:* Fermilab, U.S. Department of Energy

**Image:Heisenberg gamma ray microscope.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Heisenberg\\_gamma\\_ray\\_microscope.svg](http://en.wikipedia.org/w/index.php?title=File:Heisenberg_gamma_ray_microscope.svg) *License:* GNU Free Documentation License *Contributors:* pari

**File:Quantum tunnel effect and its application to the scanning tunneling microscope.ogv** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Quantum\\_tunnel\\_effect\\_and\\_its\\_application\\_to\\_the\\_scanning\\_tunneling\\_microscope.ogv](http://en.wikipedia.org/w/index.php?title=File:Quantum_tunnel_effect_and_its_application_to_the_scanning_tunneling_microscope.ogv) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Jubobroff

**Image:TunnelEffektKling1.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:TunnelEffektKling1.png> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Felix Kling

**File:Quantum Tunnelling animation.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Quantum\\_Tunnelling\\_animation.gif](http://en.wikipedia.org/w/index.php?title=File:Quantum_Tunnelling_animation.gif) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Yuvalr

**Image:EffetTunnel.gif** *Source:* <http://en.wikipedia.org/w/index.php?title=File:EffetTunnel.gif> *License:* GNU Free Documentation License *Contributors:* Original uploader was Jean-Christophe BENOIST at fr.wikipedia

**File:Wigner function for tunnelling.ogv** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Wigner\\_function\\_for\\_tunnelling.ogv](http://en.wikipedia.org/w/index.php?title=File:Wigner_function_for_tunnelling.ogv) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Denysbondar

**File:Rtd seq v3.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Rtd\\_seq\\_v3.gif](http://en.wikipedia.org/w/index.php?title=File:Rtd_seq_v3.gif) *License:* Creative Commons Attribution 3.0 *Contributors:* Saumitra R Mehrotra & Gerhard Klimeck

**Image:Hydrogen.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Hydrogen.svg> *License:* GNU Free Documentation License *Contributors:* Mets501, Mion, Soeb, Treisijs, Xxxx00, 6 anonymous edits

**Image:Holography-reconstruct.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Holography-reconstruct.png> *License:* GNU Free Documentation License *Contributors:* Original uploader was DrBob at en.wikipedia

**Image:Hersenen.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Hersenen.png> *License:* Creative Commons Attribution-Sharealike 3.0,2.5,2.0,1.0 *Contributors:* Bertyhell

**File:Eta Carinae Nebula 1.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Eta\\_Carinae\\_Nebula\\_1.jpg](http://en.wikipedia.org/w/index.php?title=File:Eta_Carinae_Nebula_1.jpg) *License:* Public domain *Contributors:* Hubble image: NASA, ESA, N. Smith (University of California, Berkeley), and The Hubble Heritage Team (STScI/AURA) CTIO data: N. Smith (University of California, Berkeley) and NOAO/AURA/NSF

**File:Leibniz.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Leibniz.jpg> *License:* Public Domain *Contributors:* anonimus

**File:Gustave coriolis.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Gustave\\_coriolis.jpg](http://en.wikipedia.org/w/index.php?title=File:Gustave_coriolis.jpg) *License:* Public Domain *Contributors:* Kilom691, Malo, Mschindwein, Mu

**File:SS-joule.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:SS-joule.jpg> *License:* Public Domain *Contributors:* Paddy, Serinde, 1 anonymous edits

**File:Joule's Apparatus (Harper's Scan).png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Joule's\\_Apparatus\\_\(Harper's\\_Scan\).png](http://en.wikipedia.org/w/index.php?title=File:Joule's_Apparatus_(Harper's_Scan).png) *License:* Public Domain *Contributors:* Ariadacapo, Chowbok, Pieter Kuiper, 1 anonymous edits

**File:Casimir plates.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Casimir\\_plates.svg](http://en.wikipedia.org/w/index.php?title=File:Casimir_plates.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Emok

**File:Casimir plates bubbles.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Casimir\\_plates\\_bubbles.svg](http://en.wikipedia.org/w/index.php?title=File:Casimir_plates_bubbles.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Emok, MissMJ

**File:Water wave analogue of Casimir effect.ogv** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Water\\_wave\\_analogue\\_of\\_Casimir\\_effect.ogv](http://en.wikipedia.org/w/index.php?title=File:Water_wave_analogue_of_Casimir_effect.ogv) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Denysbondar

**Image:RobertFuddBewusstsein17Jh.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:RobertFuddBewusstsein17Jh.png> *License:* Public Domain *Contributors:* Robert Fludd

**Image:Descartes mind and body.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Descartes\\_mind\\_and\\_body.gif](http://en.wikipedia.org/w/index.php?title=File:Descartes_mind_and_body.gif) *License:* Public Domain *Contributors:* Hsarrazin, Magog the Ogre, Mdd, Polarlys, Tomia, Was a bee

**File:Necker cube.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Necker\\_cube.svg](http://en.wikipedia.org/w/index.php?title=File:Necker_cube.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* BenFrantzDale

**File:Neural Correlates Of Consciousness.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Neural\\_Correlates\\_Of\\_Consciousness.jpg](http://en.wikipedia.org/w/index.php?title=File:Neural_Correlates_Of_Consciousness.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0,2.5,2.0,1.0 *Contributors:* Christof Koch

**File:Abbot of Watkungtaphao in Phu Soidao Waterfall.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Abbot\\_of\\_Watkungtaphao\\_in\\_Phu\\_Soidao\\_Waterfall.jpg](http://en.wikipedia.org/w/index.php?title=File:Abbot_of_Watkungtaphao_in_Phu_Soidao_Waterfall.jpg) *License:* Creative Commons Attribution 3.0 *Contributors:* Beta M, Tevaprapas, Tmd

**File:Neon colour spreading illusion no caption.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Neon\\_colour\\_spreading\\_illusion\\_no\\_caption.png](http://en.wikipedia.org/w/index.php?title=File:Neon_colour_spreading_illusion_no_caption.png) *License:* GNU Free Documentation License *Contributors:* Original uploader was RobinH at en.wikibooks, modified by The Anome at en.wikipedia

**File:JohnLocke.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:JohnLocke.png> *License:* Public Domain *Contributors:* Dcoetzee, Diomede, FranksValli, Herbythyme, Hystrix, Kilom691, Raymond, Shakko, 7 anonymous edits

**File:Stuart Hameroff TASC2008.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Stuart\\_Hameroff\\_TASC2008.JPG](http://en.wikipedia.org/w/index.php?title=File:Stuart_Hameroff_TASC2008.JPG) *License:* Creative Commons Attribution 3.0 *Contributors:* Zereskh

**File:Roger Penrose-6Nov2005.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Roger\\_Penrose-6Nov2005.jpg](http://en.wikipedia.org/w/index.php?title=File:Roger_Penrose-6Nov2005.jpg) *License:* Creative Commons Attribution-ShareAlike 2.0 *Contributors:* Festival della Scienza from Genova

**Image:Roger-Penrose-Kachelstruktur.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Roger-Penrose-Kachelstruktur.jpg> *License:* Creative Commons Attribution-ShareAlike 2.5 *Contributors:* Avatar, Darapti, Mattes, Schimmelreiter, Sparkit, Vincent Steenberg, 3 anonymous edits

**File:Roger Penrose 9671.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Roger\\_Penrose\\_9671.JPG](http://en.wikipedia.org/w/index.php?title=File:Roger_Penrose_9671.JPG) *License:* Creative Commons Attribution 3.0 *Contributors:* Biswarup Ganguly

**File:Penrose in University of Santiago de Compostela to picked up the Fonseca prize.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Penrose\\_in\\_University\\_of\\_Santiago\\_de\\_Compostela\\_to\\_picked\\_up\\_the\\_Fonseca\\_prize.jpg](http://en.wikipedia.org/w/index.php?title=File:Penrose_in_University_of_Santiago_de_Compostela_to_picked_up_the_Fonseca_prize.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 *Contributors:* User:Spytrap

**File:RogerPenrose CapturingInfinity cropped.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:RogerPenrose\\_CapturingInfinity\\_cropped.jpg](http://en.wikipedia.org/w/index.php?title=File:RogerPenrose_CapturingInfinity_cropped.jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* RogerPenrose\_CapturingInfinity.jpg: Procsilas Moscas derivative work: Saibo (Δ)

**File:Holomouse2.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Holomouse2.jpg> *License:* Public Domain *Contributors:* Holo-Mouse.jpg: Georg-Johann Lay derivative work: Epzcaw (talk)

**File:Portrait of Yuri Denisjuk.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Portrait\\_of\\_Yuri\\_Denisjuk.jpg](http://en.wikipedia.org/w/index.php?title=File:Portrait_of_Yuri_Denisjuk.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 *Contributors:* Svetiduh

**File:Holograph-record.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Holograph-record.svg> *License:* GNU Free Documentation License *Contributors:* User:DrBob, User:SreeBot

**File:Holography-reconstruct.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Holography-reconstruct.svg> *License:* GNU Free Documentation License *Contributors:* User:DrBob, User:SreeBot

**File:Holographic recording.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Holographic\\_recording.jpg](http://en.wikipedia.org/w/index.php?title=File:Holographic_recording.jpg) *License:* Public domain *Contributors:* Epzcaw (talk)

**File:Zonenplatte Cosinus.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Zonenplatte\\_Cosinus.png](http://en.wikipedia.org/w/index.php?title=File:Zonenplatte_Cosinus.png) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Georg Wiora (Dr. Schorsch)

**File:Holography setup.jpeg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Holography\\_setup.jpeg](http://en.wikipedia.org/w/index.php?title=File:Holography_setup.jpeg) *License:* Creative Commons Attribution 3.0 *Contributors:* Epzcaw

**File:Rainbow hologram.jpeg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Rainbow\\_hologram.jpeg](http://en.wikipedia.org/w/index.php?title=File:Rainbow_hologram.jpeg) *License:* Creative Commons Attribution-ShareAlike 3.0 *Contributors:* User:Epzcaw

**File:broken hologram.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Broken\\_hologram.jpg](http://en.wikipedia.org/w/index.php?title=File:Broken_hologram.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 *Contributors:* User:Epzcaw

**File:Contest3.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Contest3.jpg> *License:* Creative Commons Attribution-ShareAlike 3.0 *Contributors:* Dabatt123 (talk)

**File:Hologramm.JPG** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Hologramm.JPG> *License:* Public Domain *Contributors:* User:Mattes

**File:Euclid.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Euclid.jpg> *License:* Public Domain *Contributors:* Cyberpunk, Deerstop, Fishbone16, HUB, Mattes, Petropoxy (Lithoderm Proxy), Saikko, 9 anonymous edits

**Image:Kapitolinischer Pythagoras adjusted.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Kapitolinischer\\_Pythagoras\\_adjusted.jpg](http://en.wikipedia.org/w/index.php?title=File:Kapitolinischer_Pythagoras_adjusted.jpg) *License:* GNU Free Documentation License *Contributors:* Original uploader was Galilea at de.wikipedia

**File:maya.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Maya.svg> *License:* GNU Free Documentation License *Contributors:* Bryan Derksen

**File:GodfreyKneller-IsaacNewton-1689.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:GodfreyKneller-IsaacNewton-1689.jpg> *License:* Public Domain *Contributors:* Algorithm, Beyond My Ken, Bjankuloski06en, Ecummenic, Grenavitar, Infrogmation, Kelson, Kilom691, Porao, Saperaud, Semnoz, Siebrand, Sparkit, Thomas Gun, Vonvon, Wikiklaas, Wknight94, Wst, Zaphod, 7 anonymous edits

**File:Leonhard Euler 2.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Leonhard\\_Euler\\_2.jpg](http://en.wikipedia.org/w/index.php?title=File:Leonhard_Euler_2.jpg) *License:* Public Domain *Contributors:* Funck77, Haham hanuka, Herbythyme, Plindenbaum, Rd232, Serge Lachinov, Shakko, Soerfm, 7 anonymous edits

**File:Abacus 6.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Abacus\\_6.png](http://en.wikipedia.org/w/index.php?title=File:Abacus_6.png) *License:* unknown *Contributors:* Flominator, German, Grön, Luestling, RHorning

**File:Venn A intersect B.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Venn\\_A\\_intersect\\_B.svg](http://en.wikipedia.org/w/index.php?title=File:Venn_A_intersect_B.svg) *License:* Public Domain *Contributors:* Cepheus, Kbolino, Mate2code, Olaf, Patricia.fidi, Ricordisamoa, Waldir

**File:Commutative diagram for morphism.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Commutative\\_diagram\\_for\\_morphism.svg](http://en.wikipedia.org/w/index.php?title=File:Commutative_diagram_for_morphism.svg) *License:* Public Domain *Contributors:* User:Cepheus

**File:DFAexample.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:DFAexample.svg> *License:* Public Domain *Contributors:* Cepheus

**File:Elliptic curve simple.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Elliptic\\_curve\\_simple.svg](http://en.wikipedia.org/w/index.php?title=File:Elliptic_curve_simple.svg) *License:* GNU Free Documentation License *Contributors:* derivative work: Pbroks13 (talk) Elliptic\_curve\_simple.png: Created by Sean κ. + 23:33, 27 May 2005 (UTC)

**File:Rubik's cube.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Rubik's\\_cube.svg](http://en.wikipedia.org/w/index.php?title=File:Rubik's_cube.svg) *License:* GNU Free Documentation License *Contributors:* This image was created by me, Booyabazooka

**File:Group diagram D6.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Group\\_diagram\\_D6.svg](http://en.wikipedia.org/w/index.php?title=File:Group_diagram_D6.svg) *License:* Public Domain *Contributors:* User:Cepheus

**File:Lattice of the divisibility of 60.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Lattice\\_of\\_the\\_divisibility\\_of\\_60.svg](http://en.wikipedia.org/w/index.php?title=File:Lattice_of_the_divisibility_of_60.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* User:Ed g2s

**File:Braid-modular-group-cover.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Braid-modular-group-cover.svg> *License:* Public Domain *Contributors:* Nils R. Barth

**File:Illustration to Euclid's proof of the Pythagorean theorem.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Illustration\\_to\\_Euclid's\\_proof\\_of\\_the\\_Pythagorean\\_theorem.svg](http://en.wikipedia.org/w/index.php?title=File:Illustration_to_Euclid's_proof_of_the_Pythagorean_theorem.svg) *License:* Public Domain *Contributors:* Darapti, Gerbrant, 2 anonymous edits

**File:Sinusvåg 400px.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Sinusvåg\\_400px.png](http://en.wikipedia.org/w/index.php?title=File:Sinusvåg_400px.png) *License:* unknown *Contributors:* User Solkoll on sv.wikipedia

**File:Hyperbolic triangle.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Hyperbolic\\_triangle.svg](http://en.wikipedia.org/w/index.php?title=File:Hyperbolic_triangle.svg) *License:* Public Domain *Contributors:* Bender235, LucasVB, 1 anonymous edits

**File:Torus.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Torus.png> *License:* Public Domain *Contributors:* LucasVB, Rimshot, SharkD

**File:Mandel zoom 07 satellite.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Mandel\\_zoom\\_07\\_satellite.jpg](http://en.wikipedia.org/w/index.php?title=File:Mandel_zoom_07_satellite.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* User:Wolfgangbeyber

**File:Measure illustration.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Measure\\_illustration.png](http://en.wikipedia.org/w/index.php?title=File:Measure_illustration.png) *License:* Public Domain *Contributors:* Oleg Alexandrov

**File:Integral as region under curve.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Integral\\_as\\_region\\_under\\_curve.svg](http://en.wikipedia.org/w/index.php?title=File:Integral_as_region_under_curve.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* 4C

**File:Vector field.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Vector\\_field.svg](http://en.wikipedia.org/w/index.php?title=File:Vector_field.svg) *License:* Public Domain *Contributors:* Fibonacci.

**File:Airflow-Obstructed-Duct.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Airflow-Obstructed-Duct.png> *License:* Public Domain *Contributors:* Original uploader was User A1 at en.wikipedia

**File:Limitcycle.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Limitcycle.svg> *License:* Creative Commons Attribution-ShareAlike 3.0 *Contributors:* User:Gargan

**File:Lorenz attractor.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Lorenz\\_attractor.svg](http://en.wikipedia.org/w/index.php?title=File:Lorenz_attractor.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* User:Dschwen

**File:Conformal grid after Möbius transformation.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Conformal\\_grid\\_after\\_Möbius\\_transformation.svg](http://en.wikipedia.org/w/index.php?title=File:Conformal_grid_after_Möbius_transformation.svg) *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* Lokal\_Profil

**Image:Gravitation space source.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Gravitation\\_space\\_source.png](http://en.wikipedia.org/w/index.php?title=File:Gravitation_space_source.png) *License:* GNU Free Documentation License *Contributors:* Adam majewski, Duesentrieb, Schekinov Alexey Victorovich, Superborsuk, WikipediaMaster

**Image:BernoullisLawDerivationDiagram.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:BernoullisLawDerivationDiagram.svg> *License:* GNU Free Documentation License *Contributors:* MannyMax (original)

**Image:Composite trapezoidal rule illustration small.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Composite\\_trapezoidal\\_rule\\_illustration\\_small.svg](http://en.wikipedia.org/w/index.php?title=File:Composite_trapezoidal_rule_illustration_small.svg) *License:* Attribution *Contributors:* derivative work: Pbroks13 (talk) Composite\_trapezoidal\_rule\_illustration\_small.png: en>User:Oleg Alexandrov

**Image:Maximum boxed.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Maximum\\_boxed.png](http://en.wikipedia.org/w/index.php?title=File:Maximum_boxed.png) *License:* Public Domain *Contributors:* Freiddy

**Image:Two red dice 01.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Two\\_red\\_dice\\_01.svg](http://en.wikipedia.org/w/index.php?title=File:Two_red_dice_01.svg) *License:* Public Domain *Contributors:* Stephen Silver

**Image:Oldfaithful3.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Oldfaithful3.png> *License:* Public Domain *Contributors:* Anynobody, Cwbn (commons), Maksim, Mdd, Nandhp, Oleg Alexandrov, WikipediaMaster, 6 anonymous edits

**File:Caesar3.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Caesar3.svg> *License:* Public Domain *Contributors:* Cepheus

**Image:Market Data Index NYA on 20050726 202628 UTC.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Market\\_Data\\_Index\\_NYA\\_on\\_20050726\\_202628\\_UTC.png](http://en.wikipedia.org/w/index.php?title=File:Market_Data_Index_NYA_on_20050726_202628_UTC.png) *License:* Public Domain *Contributors:* Denniss, Jodo, Tangopaso

**Image:Arbitrary-gametrete-solved.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Arbitrary-gametrete-solved.svg> *License:* Public Domain *Contributors:* derivative work: Qef (talk) Arbitrary-gametrete-solved.png: Original uploader was Fieari at en.wikipedia, this is a hand-crafted SVG alternative.

**Image:Signal transduction pathways.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Signal\\_transduction\\_pathways.svg](http://en.wikipedia.org/w/index.php?title=File:Signal_transduction_pathways.svg) *License:* GNU Free Documentation License *Contributors:* cybertory

**Image:Ch4-structure.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Ch4-structure.png> *License:* GNU Free Documentation License *Contributors:* Benjah-bmm27, Cwbn (commons), Dbc334, Maksim, Razorbliss, 1 anonymous edits

**Image:GDP PPP Per Capita IMF 2008.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:GDP\\_PPP\\_Per\\_Capita\\_IMF\\_2008.png](http://en.wikipedia.org/w/index.php?title=File:GDP_PPP_Per_Capita_IMF_2008.png) *License:* Creative Commons Attribution 3.0 *Contributors:* Sbw01f

**Image:Simple feedback control loop2.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Simple\\_feedback\\_control\\_loop2.svg](http://en.wikipedia.org/w/index.php?title=File:Simple_feedback_control_loop2.svg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Simple\_feedback\_control\_loop2.png: Corona derivative work: Hazmat2 (talk)

**File:Carl Friedrich Gauss.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Carl\\_Friedrich\\_Gauss.jpg](http://en.wikipedia.org/w/index.php?title=File:Carl_Friedrich_Gauss.jpg) *License:* Public Domain *Contributors:* Gottlieb BiermannA, Wittmann (photo)

**Image:Tepantitla mural, Ballplayer A (Daquella manera).jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Tepantitla\\_mural\\_Ballplayer\\_A\\_\(Daquella\\_manera\).jpg](http://en.wikipedia.org/w/index.php?title=File:Tepantitla_mural_Ballplayer_A_(Daquella_manera).jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* Dodo, FlickreviewR, Madman2001

**Image:Cuneiform script2.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Cuneiform\\_script2.png](http://en.wikipedia.org/w/index.php?title=File:Cuneiform_script2.png) *License:* Public Domain *Contributors:* Cuneiform\_script2.jpg: derivative work: Yjenith (talk)

**Image:Girls learning sign language.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Girls\\_learning\\_sign\\_language.jpg](http://en.wikipedia.org/w/index.php?title=File:Girls_learning_sign_language.jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* David Fulmer from Pittsburgh

**Image:Braille house09.JPG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Braille\\_house09.JPG](http://en.wikipedia.org/w/index.php?title=File:Braille_house09.JPG) *License:* Creative Commons Attribution-Sharealike 3.0,2.5,2.0,1.0 *Contributors:* Kou07kou

**File:ASL family.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:ASL\\_family.jpg](http://en.wikipedia.org/w/index.php?title=File:ASL_family.jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* David Fulmer from Pittsburgh

**Image:BBC-artefacts.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:BBC-artefacts.jpg> *License:* GNU Free Documentation License *Contributors:* Original uploader was Chenshilwood at en.wikipedia

**Image:Pieter Bruegel the Elder - The Tower of Babel (Vienna) - Google Art Project - edited.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Pieter\\_Bruegel\\_the\\_Elder\\_-\\_The\\_Tower\\_of\\_Babel\\_\(Vienna\)\\_-\\_Google\\_Art\\_Project\\_-\\_edited.jpg](http://en.wikipedia.org/w/index.php?title=File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-_Google_Art_Project_-_edited.jpg) *License:* unknown *Contributors:* Dcoetzee, Foundling, Hahc21, Julia W

**Image:Sir William Jones.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Sir\\_William\\_Jones.jpg](http://en.wikipedia.org/w/index.php?title=File:Sir_William_Jones.jpg) *License:* Public Domain *Contributors:* BishkekRocks, Cirt, Ecummenic, Hsarrazin, Jonsafari, Lokal Profil, Mattes, Mu, Shakko, Thomas Gun, VladiMens

**Image:Ferdinand de Saussure by Julien.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Ferdinand\\_de\\_Saussure\\_by\\_Julien.png](http://en.wikipedia.org/w/index.php?title=File:Ferdinand_de_Saussure_by_Julien.png) *License:* Public Domain *Contributors:* "F. Jullien Genève", maybe Frank-Henri Jullien (1882–1938)

**Image:Noam chomsky cropped.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Noam\\_chomsky\\_cropped.jpg](http://en.wikipedia.org/w/index.php?title=File:Noam_chomsky_cropped.jpg) *License:* Copyrighted free use *Contributors:* Created by John Soares, uploaded to Commons by Stevertigo, then modified by Verdy p.

**Image:Brain Surface Gyri.SVG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Brain\\_Surface\\_Gyri.SVG](http://en.wikipedia.org/w/index.php?title=File:Brain_Surface_Gyri.SVG) *License:* Creative Commons Attribution-Share Alike *Contributors:* James.mcd.nz

**Image:Illu01 head neck.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Illu01\\_head\\_neck.jpg](http://en.wikipedia.org/w/index.php?title=File:Illu01_head_neck.jpg) *License:* Public Domain *Contributors:* Arcadian

**Image:Spectrogram -iua-.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Spectrogram\\_-iua-.png](http://en.wikipedia.org/w/index.php?title=File:Spectrogram_-iua-.png) *License:* Creative Commons Attribution 2.0 *Contributors:* Ishwar, Mike.lifeguard, Moyogo, 2 anonymous edits

**Image:Real-time MRI - Speaking (Chinese).ogv** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Real-time\\_MRI\\_-\\_Speaking\\_\(Chinese\).ogv](http://en.wikipedia.org/w/index.php?title=File:Real-time_MRI_-_Speaking_(Chinese).ogv) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* Martin.uecker

**Image:Ancient Tamil Script.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Ancient\\_Tamil\\_Script.jpg](http://en.wikipedia.org/w/index.php?title=File:Ancient_Tamil_Script.jpg) *License:* Creative Commons Attribution 2.0 *Contributors:* Symphony Symphony from New York, US

**Image:ManSpec.png** *Source:* <http://en.wikipedia.org/w/index.php?title=File:ManSpec.png> *License:* Creative Commons Zero *Contributors:* Grechuk

**Image:Hangul wi.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Hangul\\_wi.svg](http://en.wikipedia.org/w/index.php?title=File:Hangul_wi.svg) *License:* unknown *Contributors:* Kjoonlee, Kwj2772, Mechamind90, Sarang, Waldir

**Image:KSL wi.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:KSL\\_wi.jpg](http://en.wikipedia.org/w/index.php?title=File:KSL_wi.jpg) *License:* GNU Free Documentation License *Contributors:* Original uploader was Kwamikagami at en.wikipedia

**File:Basic constituent structure analysis English sentence.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Basic\\_constituent\\_structure\\_analysis\\_English\\_sentence.svg](http://en.wikipedia.org/w/index.php?title=File:Basic_constituent_structure_analysis_English_sentence.svg) *License:* Public Domain *Contributors:* Adam majewski, AnonMoos

**Image:Speech Client WEB (5).jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Speech\\_Client\\_WEB\\_\(5\).jpg](http://en.wikipedia.org/w/index.php?title=File:Speech_Client_WEB_(5).jpg) *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Madmike1952

**Image:Lakhovsky Conversation.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Lakhovsky\\_Conversation.jpg](http://en.wikipedia.org/w/index.php?title=File:Lakhovsky_Conversation.jpg) *License:* Public Domain *Contributors:* AndreasPraefcke, Bukk, Dia^, Mattes, Peter Isotalo, Ticketautomat, 1 anonymous edits

**Image:Winnipeg Forks - Plains Cree Inscription.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Winnipeg\\_Forks\\_-\\_Plains\\_Cree\\_Inscription.jpg](http://en.wikipedia.org/w/index.php?title=File:Winnipeg_Forks_-_Plains_Cree_Inscription.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Ardfern, Diderot, G.dallorto, Man vyi, Skeezi1000, Tar-ba-gan, 2 anonymous edits

**Image:Beowulf.firstpage.jpeg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Beowulf.firstpage.jpeg> *License:* Public Domain *Contributors:* Andrew Gray, Caravaca, Davepape, EugeneZelenko, Evrik, Jastrow, Jonathunder, Leos vän, Neddyseagoon, Ranveig, Semnoz, Verica Atrebatum, 3 anonymous edits

**File:Human Language Families Map.PNG** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Human\\_Language\\_Families\\_Map.PNG](http://en.wikipedia.org/w/index.php?title=File:Human_Language_Families_Map.PNG) *License:* GNU Free Documentation License *Contributors:* . Later version(s) were uploaded by .

**File:Linguistic diversity.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Linguistic\\_diversity.png](http://en.wikipedia.org/w/index.php?title=File:Linguistic_diversity.png) *License:* Public Domain *Contributors:* Davius

**Image:Sandro Botticelli - La Carte de l'Enfer.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Sandro\\_Botticelli\\_-\\_La\\_Carte\\_de\\_l'Enfer.jpg](http://en.wikipedia.org/w/index.php?title=File:Sandro_Botticelli_-_La_Carte_de_l'Enfer.jpg) *License:* Public Domain *Contributors:* Enzian44, Mattes, Mel22, Oxxo

**File:Schumann resonance animation.ogv** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Schumann\\_resonance\\_animation.ogv](http://en.wikipedia.org/w/index.php?title=File:Schumann_resonance_animation.ogv) *License:* Public Domain *Contributors:* NASA/Goddard Space Flight Center

**File:Lightning3.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:Lightning3.jpg> *License:* Public Domain *Contributors:* U.S. Air Force photo by Edward Aspera Jr.

**File:NeTube.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:NeTube.jpg> *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* User:Pslawinski

**File:Plasma-lamp 2.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Plasma-lamp\\_2.jpg](http://en.wikipedia.org/w/index.php?title=File:Plasma-lamp_2.jpg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Luc Viatour

**File:Space Shuttle Atlantis in the sky on July 21, 2011, to its final landing.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Space\\_Shuttle\\_Atlantis\\_in\\_the\\_sky\\_on\\_July\\_21,\\_2011,\\_to\\_its\\_final\\_landing.jpg](http://en.wikipedia.org/w/index.php?title=File:Space_Shuttle_Atlantis_in_the_sky_on_July_21,_2011,_to_its_final_landing.jpg) *License:* Public Domain *Contributors:* Huntster, Mattisse

**File:plasma fountain.gif** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Plasma\\_fountain.gif](http://en.wikipedia.org/w/index.php?title=File:Plasma_fountain.gif) *License:* Public Domain *Contributors:* Original uploader was Iantresman at en.wikipedia

**File:Plasma scaling.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Plasma\\_scaling.svg](http://en.wikipedia.org/w/index.php?title=File:Plasma_scaling.svg) *License:* Public domain *Contributors:* Croquant, Cwbn (commons), Magog the Ogre, Pierre cb, Pieter Kuiper, WikipediaMaster

**File:Lightning over Oradea Romania 3.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Lightning\\_over\\_Oradea\\_Romania\\_3.jpg](http://en.wikipedia.org/w/index.php?title=File:Lightning_over_Oradea_Romania_3.jpg) *License:* Public Domain *Contributors:* Mircea Madau

**File:Main tycho remnant full.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Main\\_tycho\\_remnant\\_full.jpg](http://en.wikipedia.org/w/index.php?title=File:Main_tycho_remnant_full.jpg) *License:* Public Domain *Contributors:* NASA/MPIA/Calar Alto Observatory, Oliver Krause et al.

**File:Magnetic rope.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Magnetic\\_rope.svg](http://en.wikipedia.org/w/index.php?title=File:Magnetic_rope.svg) *License:* Public Domain *Contributors:* CharlesChandler

**File:Plasma jacobs ladder.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Plasma\\_jacobs\\_ladder.jpg](http://en.wikipedia.org/w/index.php?title=File:Plasma_jacobs_ladder.jpg) *License:* Creative Commons Attribution-Sharelike 3.0 *Contributors:* Chocolateak

**File:Simple representation of a discharge tube - plasma.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Simple\\_representation\\_of\\_a\\_discharge\\_tube\\_-\\_plasma.png](http://en.wikipedia.org/w/index.php?title=File:Simple_representation_of_a_discharge_tube_-_plasma.png) *License:* Creative Commons Attribution-Sharelike 3.0 *Contributors:* Rudolfensis

**File:Cascade process of ionization.png** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Cascade\\_process\\_of\\_ionization.png](http://en.wikipedia.org/w/index.php?title=File:Cascade_process_of_ionization.png) *License:* Creative Commons Attribution-Sharelike 3.0 *Contributors:* Rudolfensis

**File:HallThruster 2.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:HallThruster\\_2.jpg](http://en.wikipedia.org/w/index.php?title=File:HallThruster_2.jpg) *License:* Public Domain *Contributors:* Original uploader was Dstaack at en.wikipedia

**File:Wispy 'Plasma Dancer' on the limb of the Sun.ogv** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Wispy\\_'Plasma\\_Dancer'\\_on\\_the\\_limb\\_of\\_the\\_Sun.ogv](http://en.wikipedia.org/w/index.php?title=File:Wispy_'Plasma_Dancer'_on_the_limb_of_the_Sun.ogv) *License:* Public Domain *Contributors:* NASA/Goddard Space Flight Center

**File:Plasma Spraying Process.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Plasma\\_Spraying\\_Process.jpg](http://en.wikipedia.org/w/index.php?title=File:Plasma_Spraying_Process.jpg) *License:* Attribution *Contributors:* <http://www.zircotec.com>

**Image:Gottfried Wilhelm von Leibniz.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Gottfried\\_Wilhelm\\_von\\_Leibniz.jpg](http://en.wikipedia.org/w/index.php?title=File:Gottfried_Wilhelm_von_Leibniz.jpg) *License:* Public Domain *Contributors:* AndreasPraefcke, Auntof6, Beria, Beyond My Ken, Boo-Boo Baroo, Cirt, Davidlud, Ecummenic, Eusebius, Factumquintus, FalconL, Gabor, Luestling, Mattes, Schaengel89, Shakko, Svencb, Tomisti, 5 anonymous edits

**Image:Godfrey Kneller-IsaacNewton-1689.jpg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:GodfreyKneller-IsaacNewton-1689.jpg> *License:* Public Domain *Contributors:* Algorithme, Beyond My Ken, Bjankuloski06en, Ecummenic, Grenavitar, Infrogmation, Kelson, Kilom691, Porao, Saperaud, Semnoz, Siebrand, Sparkit, Thomas Gun, Vonvon, Wikiklaas, Wknight94, Wst, Zaphod, 7 anonymous edits

**Image:Immanuel Kant (painted portrait).jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Immanuel\\_Kant\\_\(painted\\_portrait\).jpg](http://en.wikipedia.org/w/index.php?title=File:Immanuel_Kant_(painted_portrait).jpg) *License:* Public Domain *Contributors:* unspecified

**Image:Sphere closed path.svg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Sphere\\_closed\\_path.svg](http://en.wikipedia.org/w/index.php?title=File:Sphere_closed_path.svg) *License:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors:* Luca Antonelli (Luke Antony)

**Image:Carl Friedrich Gauss.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Carl\\_Friedrich\\_Gauss.jpg](http://en.wikipedia.org/w/index.php?title=File:Carl_Friedrich_Gauss.jpg) *License:* Public Domain *Contributors:* Gottlieb BiermannA, Wittmann (photo)

**Image:Young Poincare.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Young\\_Poincare.jpg](http://en.wikipedia.org/w/index.php?title=File:Young_Poincare.jpg) *License:* Public Domain *Contributors:* Badzil, Christian1985, Clindberg, Hohum, Homonihilis, Materialscientist, Soerfm

**Image:Albert Einstein Head.jpg** *Source:* [http://en.wikipedia.org/w/index.php?title=File:Albert\\_Einstein\\_Head.jpg](http://en.wikipedia.org/w/index.php?title=File:Albert_Einstein_Head.jpg) *License:* Public Domain *Contributors:* Photograph by Oren Jack Turner, Princeton, N.J. Modified with Photoshop by PM\_Poon and later by Dantadd.

# License

---

Creative Commons Attribution-Share Alike 3.0 Unported  
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)

---