A

ABDUCTION

Abduction is a form of reasoning distinct from deductive and inductive reasoning. It has often been referred to as the *commonsense* approach to reasoning because it is used in everyday life as well as in scientific argumentation. It has also been called *inference to the best explanation* because people reason from something that has been observed to an explanation for it. The philosopher Charles Sanders Peirce (1839–1914) first described this form of reasoning as *guessing* because he noted that you must leap to a possible explanation without knowing with certainty whether or not it is true. Abductive reasoning is a critical notion in philosophy, including the philosophy of science, as well as that of history, law, and many other disciplines. It is also a basic function of the human mind in everyday life.

Abduction in Everyday Life

Abductive reasoning is used in daily life as a commonsense approach to explaining anomalies. One of the most common examples given is that of walking out of a house to find that the grass in the lawn is wet. Because the grass is wet, one may infer that it had recently rained. If it had rained, then the fact that the grass is wet would be nothing out of the ordinary. That it had rained explains the fact that the grass is wet.

Another such case used to illustrate the process of abduction is that of a faulty light fixture. If one comes home and flips a light switch to *on*, but the light does not turn on, then one may infer that the power is out or the breaker switch has been flipped. Suppose that, although the light is still off, one sees that the clock on the microwave is on and appears to be functioning correctly. If the power were off in the whole house, then this contradiction would not occur because the microwave clock would be off as well. Now, given this new evidence, one may infer a simpler explanation to account for the light being off, namely that the light bulb has burned out and needs to be changed. In this way, abduction is the process of inferring explanations to account for the evidence or facts that are observed. The explanation can never be proven but rather only inferred.

As a generic tool of human reasoning, abduction has been of interest to philosophers because, unlike deductive reasoning, abduction is not a formalized system of proofs. In fact, the subjectivity required to decide which explanation is best has been one of the difficulties in applying abduction to science and other domains.

Abduction in Use

Abduction is commonly used in philosophical inquiry, in belief revision, in academic and professional disciplines such as medicine and anthropology, and in work-related reasoning. Philosophers of science also draw on abduction as *inference to the best explanation* to be one of the primary tools in scientific realism. Abduction is a component part of human reasoning. There have been attempts to formalize abductive reasoning to meet the needs of particular jobs or sectors of society. Some decisions require a type of validation, or rules to decide which possible explanation for an observed effect is best. On the most basic level, abduction has no such rules. It is minimalistic in this aspect and allows for subjectivity in determining the best explanation, often leading to revising previously held beliefs. For the hypothesis to be a useful one, it should be explanatory, testable, and economical. The simplest explanation is usually expected to be the best, though the subjective view can never escape the possibility for error. Below are several areas of application, and many of the uses of abduction take place among practitioners or clinicians within a particular field.

Logical Reasoning

Logical reasoning has often been split into *deductive* and *inductive* reasoning. The former begins with reasoning from a general premise to an entailed, or necessary, conclusion. There is room for error if the formal rules of logic have been followed. The latter, on the other hand, is reasoning from specific evidence or premises to a probable conclusion. The rules are not formally set out, but evidence can be collected until it sufficiently can support a conclusion.

Abductive reasoning, according to Peirce, provides a third type of reasoning by which one reasons from a conclusion to a possible explanation. This presents some particular difficulties in the realm of formal logic. Formal logic would propose that the process of inferring is deductive, as in the instance that follows:

All bachelors are *male*.

John is a bachelor.

Therefore, *John is male*.

Formally, this is written as follows where A is the antecedent and C is the consequent:

$$A \rightarrow C$$

$$A$$

$$\overline{}$$

$$\therefore C$$

In this example of *modus ponens*, we see that given the two premises (A→C and A), we must conclude that the conclusion (C) is valid. Whether or not it is true is not a concern. Logical validity is established by adhering to certain rules of deduction. Reasoning, then, happens from the given premise to infer a necessary conclusion. Abduction, however, operates in the reverse.

Abduction begins with a given conclusion or consequent, C and then seeks to infer an explanation for C. This is the formal logical fallacy of affirming the consequent because there is no guarantee that by knowing C one can infer A. That is, returning to the example above, that if *John* is male, there is no guarantee that John is a *bachelor.* Thus, abduction is a form of inference that violates formal logic, but is nonetheless critical to human thinking and philosophical reasoning. Formally, this would be written as follows:

$$A \rightarrow C$$

$$C$$

$$\overline{}$$

$$A \text{ “Invalid”}$$

While it is clear that abduction and deduction are quite different types of reasoning, the distinction between abduction and induction is less clear. Since inductive reasoning is the process of learning from example, some have proposed that induction is only a case of abduction. That is, inductive reasoning can be reduced to the more primal abductive reasoning. Other philosophers, who view induction as creating a theory to account for facts, have proposed that abduction is actually only a case of inductive reasoning. Peirce's triad of reasoning can be found to have overlap in actual use. Thus, some have proposed to use the terminology of *explanatory reasoning* to account for the specific process and product of abduction specifically.

Medical Sciences

The medical sciences make use of abduction, especially in clinical practice, through observing the symptoms of a patient and reasoning as to what might have caused the current condition. A physician may attempt to explain the cause of several symptoms experienced by a patient. Sometimes, the coexistence of several symptoms may complicate the diagnostic process as a doctor seeks to identify the root cause. Usually, the explanation that seems to explain all of the symptoms is preferable to diagnosis of several causes, thus drawing on abductive reasoning's simplicity to pinpoint a probable cause that can then be treated. Diagnosis and clinical evaluation are examples of abduction in medicine.

Law

Legal experts and lawyers often utilize abductive reasoning. Specifically within criminal cases, a jury makes a judgment regarding whether the prosecution's case sufficiently explains the evidence. The burden, then, is on the prosecution to present a plausible abduction that a judge or jury could endorse as a true narrative of events leading to the crime. The defense, on the other hand, needs only to present an alternative abduction or cast reasonable doubt on the prosecution's explanation of the evidence, to secure the verdict of not guilty. Developing a theory that can be inferred from and explain the evidence is critical in jurisprudence.

Statistics and Applied Mathematics

Abduction has been influential in applied mathematics and inferential statistics. Bayesian inferential statistics, for example, has been a useful way to make decisions based on statistical data. Through statistical evidence, one could reject or accept a hypothesis depending on whether the probability is sufficient for a particular cause to have produced the observed effect. The level of probability represents the most likely explanation, and tells the statistician which hypothesis is most statistically likely, though not necessarily the best, explanation for a consequent. This helps to remove some of the subjectivity innate to abduction. It restricts *the best explanation* to that which is *most likely*.

Abduction also has been used to develop artificial intelligence in computer programing. Parameters for abductive processes can be set, which would allow machines to analyze data and make inferences to explain that data. By modeling artificial intelligence with abduction, machines can mimic human thinking. Abduction allows for the detection of faults in computer systems, as well as belief revision. This is contrasted with deduction, which would require preprograming all possible data so the machine can abide by the formal rules of logic.

Philosophy of Science

The influence of abduction as an area of study is perhaps most clearly seen in the philosophy of science. Since the emphasis in abductive reasoning is to infer possible explanations, it can be seen that abduction does not lead to proof or certainty. As noted in the logical fallacy above, one cannot determine the antecedent from the consequent with absolute certainty. One may only infer that the antecedent is a probable cause for the consequent because the presence of the antecedent would sufficiently explain the consequent. Abduction, then, has been foundational in the area of scientific reasoning and, more broadly, what counts as science. Biological diversity, for example, can be explained by many theories. Two of the most debated are Charles Darwin's account of natural selection and creationist accounts of a divine maker. Though either of the two could serve as reasons for biological diversity, the majority of modern-day scientists have appealed to abductive reasoning to validate Darwinian evolution as the operating theory. Since one could claim that Darwin's approach sufficiently explains biological diversity, and is more plausible than a divine maker, scientists reason that the best explanation for biological diversity is natural selection. Science, then, allows for uncertain and unobserved possible causes to be taken as operational *truth* or *fact* insofar as they explain the evidence well enough. This necessarily introduces a degree of subjectivity.

The Subjectivity of Abduction

Abduction has been the target of some critique, specifically regarding the inherent subjectivity. For example, the association with abduction as the *inference to the best explanation* would require an interpretive move as to what is considered the *best* explanation for the accepted evidence. Additionally, it has fallen under criticism because it is used as a form of knowledge discovery, not as scientific verification. Thus, abduction is sometimes seen simply as hypothesis forming, and not as a valid form of logic or hypothesis verification. In epistemology, abduction can never achieve *truth* or *proof* claims but can only propose possible explanations.

A Scientific Example of Abduction

Abduction is a fundamental concept in philosophy of science. Examples of abduction can be found frequently in scientific inferences. However, one example that serves to illustrate the point is the scientific discovery of the earth's orbit and rotation. Through observation of the sun, moon, and stars, one would be tempted to think that the earth is the center of the universe and that celestial bodies revolve around it. Ptolemy's (ca. 127–145 C.E.) geocentric view was certainly influential, and it was accepted as true since it best accounted for the evidence at hand. Surely, this is what the ancients thought and indeed how they lived their lives. They observed the phenomena that the sun rises in the east and sets in the west, and the most clear-cut explanation is that the sun revolves around the earth. Some observers sought to explain that it was not the sun that was moving, but the earth that was rotating. Philolaus, the Pythagorean (470 B.C.E.–385 B.C.E.), for example, attempted to account for the same observance by theorizing that there was actually a non-earth fire-center around which the sun, moon, and planets circulated. The view was quite complicated and largely dismissed. Others had thought it possible that the earth rotated, thereby giving the appearance that the sun rose and set, but these views had failed to explain all of the cosmic observations and left early scientists believing that geocentricism was the best explanation. It was not until Nicolaus Copernicus (1473–1543 C.E.) developed his heliocentric view that the seeming anomalies in the cosmos were better explained.

Prior to Copernicus, no system naturally explained retrograde planetary motion where, at some times of the year, planets would seem to move backward in their orbits, and then forward again. Ptolemy's system needed to appeal to *epicycles*, the idea that planets were on smaller orbits which were themselves on larger orbits around the earth, to explain the qualitative features of retrograde motion. Thus, the models presented were highly complicated and made little theoretical sense though they were inferred from the observational evidence. When Copernicus developed his version of heliocentrism, he made inferences from his observations that the way to best explain the celestial movements would be if the sun was in the center and all of the planets revolved around it. This would make some planets appear to move backward because the earth would catch up to them in the orbit and bypass them. Copernicus' system also used epicycles, in fact just as many as did Ptolemy's, but only for *quantitative* features of things. The explanation for the observations was then simplified. Surely, the claim that the sun was the center of the solar system and that the earth revolved around it and rotated on its axis was not without its challengers. But through further scientific validation it was commonly accepted, verified, and improved upon. What began with observation of the sky ended with an inference as to what could possibly account for these observations. No one could actually sit on the knees of God, so to speak, and see the universe's workings, but from one perspective the workings of the universe could be inferred. Thus, science, as well as philosophy, relies on the inference to the best explanation, or abduction, to understand what may cause observed effects.

Joseph C. Rumenapp

See also Deduction; Explanation; Fact Versus Theory; Induction; Inference; Inference to the Best Explanation; Theory Construction; Values in Science

Further Readings

Aliseda, A. (2006). *Abductive reasoning: Logical investigations into discovery and explanation.* Springer.

Douven, I. (2011, Spring). Abduction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* http://plato.stanford.edu/archives/spr2011/entries/abduction/

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 74*(1), 88–95.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). Routledge.

Peirce, C. S. (1931–1935; 1958). *Collected papers of Charles Sanders Peirce* (Vols. 1–6, C. Hartshorne & P. Weiss (Eds.); Vols. 7–8 A. W. Burks (Ed.)). Harvard University Press.

Abstract Knowledge

The term *abstract* originates from the Latin *abstractus*, from the prefix *ab*-, meaning "from," "of"; and "*trahere*," which means "pull," "drag away." It refers to the mode of thinking that allows for the isolation of elements. The topic of abstract thinking can be examined from a philosophical, psychological, and even a neurophysiological standpoint. It generally assumes that the individual is capable of mentally deconstructing a whole into its parts, in order to analyze a situation simultaneously from different angles. In this way, the practice of abstract thought leads to abstract knowledge, which builds upon the individual's ability to find common properties among elements or ideas, to plan and assume hypothetically, and to think and act symbolically.

Abstract Thought and the Human World

Among the disciplines that study abstract knowledge are biology-related sciences. Ever since Charles Darwin's (1809–1882) theories of natural selection emerged, scientists have tried to identify which physical elements differentiate human beings, and their thinking processes, from those of other animals. Theorists believe that as humans created increasingly complex social models, these led to more conceptually sophisticated forms of thinking and knowledge. In order to survive and thrive, for example, humans had to develop ways of understanding the world from multiple perspectives, to draw inferences about social patterns, and regulate their own thinking and emotions according to normative group standards. Understanding and incorporating societal cues and norms form the trove of abstract knowledge that allows humans to navigate the social world around them. Language and culture, for example, were born from the need to cooperate and create complex thinking models in order to grow. What differentiates humans from other animals, as some scientists propose, are forms of thinking, learning, and cognition engendered by collaborative and communicative interaction. These interactions rely on information constructed from symbols and assumptions, which require the capability of abstract thought.

In time, humans developed cognitive capacities far more powerful than was necessary for a hunting-and-gathering group to cooperate and survive. Darwin believed that the capacity and use of language might have shaped the highly complex human brain. Subsequently, however, both biology and the cognitive sciences long evaded the issue. Some behaviorists perceived abstract thinking simply as normal animal communication rather than as manifestations of intelligence and language. The brain, working mechanically, was capable of self-organizing in order to save and process new input, such as words. Words, however, are inseparably tied to the concepts they represent. As such, they are part of a cognitive process that merges thinking and words into symbolic representation and meaningful combinations, so as to form a learned language. Therefore, language is considered one of the many instances of human abstract knowledge.

Psychologists have also been interested in developing theories of abstract knowledge. Abstract thought is different from concrete thinking, which is based on real, lived experiences. Early in life, individuals come to understand their daily experiences as real and as based upon interaction with concrete objects. In time, they begin to develop abstract thinking, that is, their own ideas or concepts. It is relevant to mention the work of Swiss psychologist Jean Piaget (1896–1980), who posited that abstract thinking actually develops at around 12 years of age. At this point, humans move from concrete thinking to being able to mentally explore abstract or immaterial ideas. Before that, children are more reliant on concrete thinking. An example of concrete thinking would be the case of very young children who believe that if they cover their eyes, they cannot be seen by others, because they, themselves, cannot see. Unless they have some mental impairment, however, adolescents and adults can understand abstract concepts, such as comprehending that the phrase "a bad taste in my mouth" can refer to a feeling that something is false, deceitful, or unfair, rather than the literal feeling of having a disagreeable taste in the mouth. In other words, abstract knowledge makes it is possible to comprehend metaphors and similes as symbolic representations or approximations of ideas. Not all metaphors, however, refer to abstract ideas. The phrase "Get out of here!" may have a real as well as a metaphorical meaning, the latter referring to an expression of incredulity. People can understand the difference because, inevitably, abstract knowledge is often built upon concrete concepts.

Early philosophical concepts of abstract thinking were first developed by the ancient Greek philosopher Aristotle (384–322 B.C.E.), especially in his theory of universals. Abstraction is a mental activity in which a conceptual property is detached from the whole, so that it may be reflected upon and analyzed, separated from the rest. Whereas Aristotle's teacher Plato (ca. 429–348 B.C.E.) believed that all objects and ideas had an essence or pure form, which could be perceived intuitively by humans, Aristotle posited that universal ideas exist but that these must be founded upon empirical data or observation. In other words, abstract knowledge must be built upon concrete information. In that way, the idea of *tree* comes from the experience and comparison of many trees, from which aggregate an individual may extrapolate a general concept of *tree,* which shares the characteristics of what they all have in common. It is then by comparing multiple objects that a property considered shared by all may be isolated or abstracted, and the isolated object is a universal. The individual then forms an idea of what a tree, which represents all trees, is. This becomes abstract knowledge, based upon exercising abstract thinking. Abstract thinking is not necessarily based on concrete or material experiences but on the capacity of making projections and building conjectures on a purely abstract and intangible plane. In this sense, many also call it *theoretical*, considered separate from material or concrete existence. One example of abstract thinking, for example, is developed in mathematics and physics, as well as in many philosophical concepts that deal with intangible ideas. Others have compared abstract thinking with the metaphorical plane, which implies the ability to represent ideas symbolically.

It is important to highlight that mental representations of real objects are aided by language. Abstract knowledge is based on formal schemas, which are thought units by which knowledge is represented. Schemas make it possible for people to have foresight and think in terms of probabilities. They also allow the people to categorize and integrate new information. Abstract thinking is fundamental to all human beings because it allows the capacity of induction and deduction, to synthesize information, and to extrapolate what is important to learn from all situations, as well as to compare, arrive at conclusions, and engage in critical thinking. In short, abstract thinking builds up abstract knowledge, which allows individuals to interpret the concrete phenomena they come across. One of the characteristics of higher level abstract thought is the capacity of simultaneously observing myriad details and evaluating various functions, as well as processing several problems, defining priorities, and responding to several tasks. Besides judgment, foresight, and reasoning, abstract thinking is also related to creativity, insight, and mental flexibility in general.

Abstract Thinking and Language

Abstract knowledge relies on language in order to conceptualize intangible concepts, such as honor, freedom, and justice, as opposed to more concrete items, such as table and tree. Abstract language is also necessary in order to relay the ideas behind metaphors, allegories, similes, idioms, and other figures of speech. Concrete language is used, for example, to explain more textual instructions, such as in manuals, medical protocols, or cooking recipes. Abstract language is used to explain more abstract ideas, such as calculus or equations.

Abstract language is also related to relative concepts. There are, of course, many degrees of abstraction. Intellectual development, as noted earlier, is tied to the gradual process of moving from the very concrete to extremely abstract thinking across a wide range of areas. In many ways, both the concrete and the abstract are specific to a particular area. For some experts, familiarity with an area makes the topic, however abstract, increasingly concrete. In this manner, then, abstract concepts can become relatively concrete to some while remaining highly abstract to others. For instance, a mathematician might find concepts of differential equations to be relatively concrete, although they remain abstract to others; an atom may be a concrete entity to a physicist, while it may be considered an abstract concept by nonexperts.

Language is a system of word-based communication that relies on both symbolic and conventional characteristics and which is complemented by culturally based nonverbal communication and word content. Culture is the framework that mediates social actions. In order to acquire language, humans must move from learning language to developing an internal language, also known as *thinking*. This knowledge, although abstract and internal, is mediated by culture and society. Piaget's theory of cognitive constructivism, for example, views language as an element of symbolic function in society. Language has a semiotic function, a process of articulation between figurative elements and their actions. The figurative elements of language are the signifiers and signified, what the words are and what they mean. Semiotics, a theory of language, is an example of knowledge which uses concrete and abstract ways of thinking, developed by scholars such as Charles Sanders Peirce and Ferdinand de Saussure.

Language and abstract thought, then, are intrinsically related in many ways. Although the ability of children to manage abstract concepts develops as they acquire language, spoken language is not the sole component in abstract thinking. Children who have been born deaf, for example, can understand many abstract concepts even though they may not develop speech language fluently. They can, however, express themselves symbolically by way of signed language and other forms of communication available to them. Abstract knowledge, then, is based on multiple forms of language and communication.

By thinking abstractly, individuals elaborate relationships of meanings, such as similarities, differences, and causalities, with symbolic meaning represented by way of language. There are three general types of thinking related to language: conceptual, judgment, and reasoning. *Conceptual thinking* refers to the representation of an object by way of abstraction and generalization. *Abstraction* provides the selection of the properties of that object which distinguish it from others, and *generalizing* refers to the assignation of common characteristics to all of those objects that share the same characteristics. Generalizations allow for different levels of depth and directions of thought that can be taken. For every problem, there are some specific directions and levels that may be more useful than others. A computer, however, can be programmed to generalize concepts to levels of abstraction and proliferation that presumably would exceed human capabilities, such as in chess-playing programs. Abstract thought is the skill, and abstract and concrete knowledge provide the tools that enable an individual to identify which directions and levels are more useful, or which direction is the most interesting. It is in these decisions that human individuality often shines forth and in which human imagination often beats the computer program.

Judgment allows the elaboration of premises that will determine the truth or falseness of a conclusion and reasoning allows the elaboration of analysis by establishing the relationship between one or more judgments or premises. Reasoning includes deductive and inductive thinking, in where deduction refers to conclusions that are arrived by way of going from the general to the specific, and induction goes from the specific to the general. An example of deductive thinking might be the formulation of the following: All humans are intelligent. All those present here are human. Therefore, all those present here are intelligent. An example of inductive thinking might be as follows: Copper expands with heat. Bronze expands with heat. Both are metals. Therefore, all metals expand with heat.

One of the most explicit theories relating language and thought is the theory of linguistic relativity developed by Benjamin Whorf (1897–1941). Whorf argued that the language structure used by an individual originates from their particular conception of the world. Different people see the world in a unique way and employ language accordingly. Therefore, an individual's language expresses a person's determined concept of the world. Many scholars believe that the cognition, thinking processes, and abstract knowledge of an individual are inevitably affected by the particular language structure of each person. For Whorf, the range of vocabulary that a person learns influences the range of activities that the person may conceive related to a particular area. It may also influence the range of creativity and abstract ideas that the person may reach. Whorf's hypothesis has two main versions. One states that language directly impacts thought. Another, more nuanced, states that language affects thought when a particular task depends directly on the properties of a specific language system. There is very little in the way of proof, to date, on the first version of Whorf's hypothesis. Language, however, is generally considered as one of the dominant elements in abstract thinking and knowledge. The abstracting process that is ever-present in language can be considered as representative of this intellectual activity.

Another important theorist who has written on language and abstract knowledge is the Soviet psychologist Lev Vygotsky (1896–1934). Vygotsky's theories on the relationship between thinking, knowledge, and language have had great impact in contemporary psychology, especially in the arena of evolutionary psychology. Vygotsky argues that thinking and language, high-level mental activities, have different genetic and ontogenic roots, yet have developed reciprocally. His stance is the opposite of Piaget's, who believed that both capacities, mental abstraction and language, were directly related. Vygotsky also critiqued behaviorist philosophies of language, finding that they highlighted types of conduct which were conditioned. For Vygotsky, the functions of thinking and language develop independently. He did acknowledge a close relationship between both, based on developmental mental stages, and argued that language could determine the development of thought and other cognitive capabilities, such as abstract knowledge.

Many contemporary experts agree not only that there are innate predispositions toward developing language in humans but also that some language competencies are independent from general cognition. Today, a variety of disciplines, from philosophy to the neurosciences, agree that language is a combination of many different abilities and processes. Neuroscience research, for example, posits that most of these processes may be studied independently of each other, which takes after Vygotsky's proposals; however, this relative autonomy is also based on understanding language as a composite of structures and subprocesses that enable the production of language, knowledge, and related cognitive processes.

Abstract Knowledge and Moral Sense

Many thinkers have long been concerned with the relationship between abstract knowledge and the development of a moral or ethical sense. Although specific differences between morals and ethics exist, both are concerned with principles, rules of conduct, or habits with respect to correct behavior and with the consequences of wrongful actions. Some experts have taken a biological approach. Recently, some neuroscientists located the seat of concrete thinking and morality—understood as rule-following—in the frontal lobes or prefrontal area of the brain. Others, from a psychological standpoint, examine how people's level of abstract thinking impacts how they develop morality and an ethical sense, with both understood by experts as a kind of abstract knowledge. Some explain that, with higher abstract skills, people become better at looking at the big picture and visualizing myriad and far-reaching consequences. Moreover, people better skilled at abstract thinking are more prone to considering fairness and harm as the basis of ethical norms. Recently, researchers found that when people are thinking abstractly, they are likelier to make judgments on the basis of core values that are applied repeatedly and across many contexts. A recent Yale University psychological study found that participants aligned with the idea that concern about justice and welfare are the core of moral values, particularly when engaging in higher level abstract thinking, such as considering far-reaching consequences or the *big picture* perspective. Views about harm and fairness are elements of human morality that prove more enduring across time than obedience to authority and in-group loyalties, which rely on concrete knowledge. Authority and in-group dynamics are much more contextual factors; that is, based on temporal or situational events, which makes them linked more closely to concrete thinking.

Moral foundations, however, transcend basic concepts of harm and fairness and physical brain functions and capabilities. The issue of moral knowledge and its relationship to abstract thinking has been examined from different perspectives by philosophers since the ancient Greeks. More recently, modern philosophers from Immanuel Kant (1724–1804) to Jürgen Habermas have written about abstract knowledge and critical thinking as related to concepts of the common good. Kant advanced the notion of deep abstract thinking in order to develop a personal and political ethics. Critics of this view, however, have argued that a highly abstracted view of morals ignores the concrete circumstances of events and may fall into rigid adherence to abstract principles that do not take into account nuanced social contexts and complexities. Nevertheless, Kant's work was invaluable in positing that only freely chosen actions can be moral. Humans, however, cannot freely choose any type of conduct that is dictated by laws that they cannot comprehend nor control, which contradicts the biology-based behaviorist model. Because each moral human behavior must be reasoned, in the Kantian view, it places the onus of a moral sense on abstract thinking and knowledge.

Habermas (1929–), like Kant, views morality within a framework of rational principles. Reasoning, one of the abstract modes of thinking, is of essential importance in preparing for moral action. Morality itself is the consequence of acting in a way that reflects a stable will in the midst of constantly changing events. In other words, moral action is based on reasoning and remains constant through different concrete contexts. Other modern thinkers equate ethics with mathematics, in that both are based on abstract concepts and universal principles. It is important to note, however, that applied ethics are always related to real cases, which brings individuals back to the concrete and contextual.

Most experts believe that in order to develop an ethical and socially responsible citizenry, it is necessary to cultivate not only their critical thinking skills, foresight capabilities, and moral imaginations but also the tools to act on what is ethically correct. There are concrete ways of developing these skills, such as engaging in punishment and rewards in order to foster mechanical learning of what is right and wrong. These, however, can go only so far, and often fail to provide individuals with the information, knowledge, and abstract thinking skills necessary for more profound moral decisions. It is also important to cultivate the abstract knowledge, concepts of justice and democracy, for example, that fosters the desire to be an ethical citizen of the world. Critical thinking skills, which are fundamentally based on abstract capacity, are crucial to developing an ethical education, moral will, and imagination or abstract knowledge necessary to engage in empathetic interaction with others.

Although most scholars often separate the fundamental thinking processes into concrete and abstract, they are not strictly separated. People often begin a thought process by engaging in concrete thinking skills and need concrete examples in order to understand abstract concepts. Terms such as freedom, morality, racism, hate, and others are highly abstract and can mean different things in different social contexts. It is necessary to contextualize with concrete examples so that people may understand how the concept is being used. The human mind is formed by a network of very rich and complex processes; striving to separate clearly its myriad thought forms and activities is often nearly impossible.

Developing Abstract Knowledge

Thinking strategies, from abstraction and generalization to synthesis and evaluation, serve humans in interpreting reality. Higher levels of thinking depend upon skillfully combining concrete and abstract mental processes; that is, transferring what is learned in one context to another. For example, learning the categorization of elements in a science class or the organization of an essay in an English class may be skills used in organizing information and content in a social science class. Some individuals are more skillful than others at building abstract knowledge, which may be the result of a wide range of factors. Some of them may be to natural predispositions, and extreme cases of difficulty in abstract thinking may be related to mental illness, frontal lobe injury, and other impairments. While there is no known specific treatment that will develop all forms of abstract thinking, some experts posit that it is possible to practice some exercises that may help, such as logic problems, mathematics puzzles, and other brain teasers. However, these are of limited use. The brain and mind processes are highly complex, and while an individual may be a proficient abstract thinker in one area, they may remain a very concrete thinker in others. Thus, somebody who excels at mathematics may have difficulty in literary theory. Exercises that help in one area of abstract thinking may not necessarily improve in others. Other strategies that may help with improving abstract thinking skills are memory problems, problem-solving, organization techniques, and cognition.

In some cases, people can adjust the environment to help concrete thinkers better understand information and content. Some of these include adopting concrete language, avoiding metaphorical concepts or deep levels of abstraction, and always explaining abstract concepts in relation to concrete examples. For example, a complex social hierarchy system in a different part of the world, which may seem extremely complex to a concrete thinker, can be made easier to comprehend by explaining it in terms of kin or social systems with which they are familiar. It is important to understand, however, that for a very concrete thinker, the progress achieved by these exercises may not translate to other content, academic, or social areas. It does, however, help the individual build up their repository of abstract knowledge in at least one area.

It is a common assumption, to a certain extent, that before the inception of adolescence, very young children do not have any ability to engage in abstract thinking. This is a misconception. Young children begin to develop abstract thinking skills and knowledge through their early preschool years. One of the ways in which they develop these is through pretend play, which can be taken to a high level of abstraction. For example, children may take an ordinary object to represent something entirely different. They can also engage in complex role-playing conducts which are, basically, variants of symbolic actions. Children learn how to read and to add and subtract by way of symbols, developing the skills that will help them later on to develop more abstract reading comprehension and mathematical skills, such as problem-solving, algebra, and calculus. There are many ways in which children engage in abstract knowledge building. Alphabets and numbers are, after all, symbols. They represent something other than their form, such as sounds and quantities. Children also draw doodles they take to represent something else. When told stories, they can imagine scenes and offer different takes on the storyline. All of these skills involve abstract thinking, which are the necessary stepping-stones to developing strong problem-solving and critical thinking skills later on. It will also be important for developing empathy and a moral self. Abstract skills enable individuals to place themselves in the position of other people and imagine their suffering; in other words, they help develop empathy. These are some of the reasons why experts recommend encouraging and aiding young children to develop abstract thinking skills through a variety of educational programs, games, and exercises.

Trudy Mercadal

See also Abstraction; Conceptual Analysis; Evolutionary Psychology; Knowledge, Logic, Formal and Informal; Thought Experiments, Scientific and Philosophical

Further Readings

Bickerton, D. (2014). *More than nature needs: Language, mind, evolution.* Harvard University Press.

Bjorklund, D. F. (2011). *Children's thinking.* Cengage Learning.

Brockman, J. (2013). *Thinking: The new science of decision-making, problem-solving and prediction.* Harper Perennial.

Chafee, J. (2012). *The philosopher's way: Thinking critically about profound ideas.* Pearson.

Cottrell, S. (2012). *Critical thinking skills: Developing effective analysis and argument.* Palgrave Macmillan.

Drapeau, P. (2014). *Sparking student creativity: Practical ways to promote innovative thinking and problem solving.* Association for Supervision & Curriculum Development.

Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them.* Penguin.

Kenrick, D. T., & Griskevicius, V. (2013). *The rational animal: How evolution made us smarter than we think.* Basic Books.

Martin, M. W. (2012). *Cognition.* Wiley.

Napier, J. L., & Luguri, J. B. (2013, November). Moral mind-sets: Abstract thinking increases preference for "Individualizing" over "Binding" moral foundations. *Social Psychological and Personality Science, 4*, 754–759.

Tomasello, M. (2014). *A natural history of human thinking.* Harvard University Press.

Yivisaker, M., Hibbard, M., & Feeney, T. (2006). What are concrete and abstract thinking? *LearNet*. Retrieved October 28, 2019, from http://www.projectlearnet.org/tutorials/concrete_vs_abstract_thinking.html

ABSTRACTION

Abstraction is defined as the development of cognitive signifiers that represent physical objects, actions, or constructs. The word *chair*, for example, might conjure up an image in the mind of a real-world object, a chair. The mind's chair is an abstraction while the physical chair is not, they are different in that one is an approximation, a set of rules, and the other is an entity. Abstraction is the development of a prototype, or set of rules, that may focus only on a limited set of details. Therefore, it is also integral to the understanding of any theory. We cannot discuss evolution, morality, or quantum dynamics without abstraction. Abstraction is not, however, just a mental process, it can be physical as it is in language, math, computer science, and genetics. Abstraction is sometimes defined as mapping a representation of a problem within another representation. Abstractions are meant to provide insights and clarify a series of observation or to provide a framework for future data analysis. As observations morph, based on an increasing number of documented exceptions, we modify the model little by little until eventually a new model is devised which supplants or augments the original model. This new prototype becomes an added tool with which to measure reality. This entry focuses on defining abstraction through a series of historical usages, defines several schools of thought regarding abstraction, and addresses the concept of abstraction in several commonly studied theoretical fields.

The Theory of Abstraction

It is a little circular that we must have a theory regarding abstraction, as abstraction is a process by which our minds develop theories, that is, by the organization and categorizing of subjects. The term itself is derived from the Latin word *abstrahere* or *to draw away from*. Within the definition of abstraction, we must presume that anything that we look at and interpret with our mind is an abstraction and therefore, so too is the theory to define the term *abstraction*. Abstraction is, in simplest terms, to distill what is generalizable and important about a subject. Our mind does this all the time, in processes that we have only begun to define scientifically. To clarify this concept, let us go back as early as we can in human history. Paintings on cave walls may be one of our first look at human abstraction: a man, a killing tool, and an animal painted on a cave wall. What we have presented is a prototype of the simplest elements necessary for *the hunt,* an abstraction. The three objects—man, tool, and animal as target—existed first, and from it we abstracted the hunt. It is a distillation of something general, with superfluous details taken out. We can see this in the picture, and we can readily understand it, but it does not show us reality, it shows us an approximation of reality which we call abstraction. Terminology is important when we describe the subjects of abstractions; an abstraction is said to be *conceptual* if it abstracts a concept and *objectual* if it abstracts an object. Therefore, we know that abstraction can indicate at least two types of subjects: objects and concepts. But more than just a generalization of a subject, abstraction is an attempt to define the essential in a subject, winnowing it down to only what is necessary. The following sections aim to describe each of these principles, generalization and essentials, through a survey of historical perspectives.

Language

Language is the most primal and fundamental example of symbolic abstraction. Human languages are at the root of all other complex abstract modeling. Words are the representations of subjects; concepts and objects as encountered above. They reside both in our minds and on our tongues. Linguists believe that abstraction is only possible as a consequence of acquired language, that is, verbalizations. The concept of *red* is an abstraction, not possible, for instance, without verbal ability in which all objects that are red can be divided from other objects only by this shared feature. Before language, one can argue that there were apples, but not *red* apples or *green* apples, as the feature red could not be segregated from the object by a mental representation until it could be generalized to define a specific detail. Written language in turn made visual symbols of sounds, further categorizing the detailed parts of words. Natural language constructs used in language and linguistics are a well-established and formalized form of symbolic concept and object abstraction. Work from the visual languages has suggested that abstraction occurs within this framework as well. A word or hand gesture is used as a representation of a concept. Furthermore, in linguistics, the separation of concepts into deeper categorization, described as the parts of speech, allows for even more elegant rules to be derived about the nature of language itself as well as to foster easier learning of syntax and grammar. Generally, knowledge acquisition via model construction is accomplished through analysis of verbal or written terminology; in other words, the abstraction of language, as we will see from the discussion of the historical root of abstraction described in the next section.

Ancient Abstractions

The ancient Greek philosopher Parmenides of Elea, in his one surviving work, focused on the power and human tendency toward abstraction. In the poem he talks of the *way of truth* and the *way of opinion*. Truth, he says, is timeless, unchanging, and *what-is*, but opinion, he says, is the way our sensory faculties lead to constructs, mostly false. Parmenides's *opinion* is what we call abstraction, and like his view on *opinion*, what we construct though abstraction is mostly false, or at least approximate. Later we will see how nature and then science have played a role in rectifying at least some of the falsehoods in our abstract models. The Athenians attempted to make a world that was as *real* as possible. Paul Feyerabend described the Athenian view of abstraction in five principles from Aristotle. The first is that the boundaries between reality and appearance cannot be established scientifically. There is a normative or *existential* component. Second, the normative component leads to debate as to the true nature of reality. Third, different lifestyles lead to different interpretations of reality. Fourth, science contains different traditions, and scientific viewpoints are not the only viewpoint. Indeed, many nonscientific cultures suggest that there is more than one way to define a reality that works. And finally, the fifth premise is that science is incomplete and fragmented. What this means is that to the ancient Athenians the world is highly parsed, and no unified view of reality exists. Today we just have more elaborate fragments. Reality then, to the ancients, is an abstraction seen differently depending on what details we make important.

Of note, it was the theological scholar William of Ockham, known for the principle referred to as *Occam's razor* (viz., explanations that posit fewer entities are to be preferred to explanations that posit more), who pioneered the field of *nominalism*, in which he claims that universals, generalizations, are the products of abstraction and have no extra-mental existence. Nominalism is a position that suggests that only specific objects exist but not universal entities; this stands in contrast to the position of *nihilism*, which denies that even simple objects exist, only mental images do. Finally, *realism* is in contrast to both nominalism and nihilism, in that it suggests that all things are physical, not mental. In these metaphysical world views, we are confronted with the notion of a mental and physical plane. Each philosophy, however, intimates the existence of the abstract, with varying degrees of grounding in the features of the physical world. Though greatly simplified in this discussion, abstractions are considered the properties of real objects in realism, mental interpretations in nominalism, and the only thing that exists in nihilism. Abstraction in this sense means that objects and concepts may not exist before we sense them or at least as we sense them. The idea of something from nothing has been a fundamental conflict between secular and religious scholars, but with the concept of abstraction we find the ability to take not only what is real and formulate nonexistent prototypes (i.e., universals), but also the ability to create patterns, where before there was nothing at all.

Natural and Scientific Abstraction

The behaviorist view is that abstraction only has bearing on reality in so far as it affects actions. Behaviors can be predicated by abstract thought, if we know the abstract model utilized, but it is the outcome that matters. Unlike the aforementioned philosophical viewpoints, *naturalists* believe that nothing exists outside of space-time. From this perspective, all abstractions exist within and as a part of natural phenomena; in other words, everything can be reduced to a physical account, even conceptions. Other viewpoints, such as Ockham's, which state that there are certain things that cannot be rationally examined, are not interpretable by our rule sticks and therefore not useful in theory design. Nature, however, played an important role in the development of abstract worldviews before a systematic scientific perspective came about. As mentioned previously, Aristotle believed that while many viewpoints existed, they were all required to be successful if they were to persist. Abstractions were valuable in so far as they helped to predict or guide successful human behavior. In many cases, these early abstractions were the models for morality and ethics on the cultural level. For example, the perception of the Egyptians that beer was a product of the gods led to widespread consumption, which in turn kept bacteria at bay and paid for the building of the pyramids. Abstraction, from this perspective, is a view of the world that allows for better control of it.

Another natural view of abstraction comes from the art world, where abstraction has been emphasized as an inspiration for all artistic forms. Abstract art is not, as the uninformed may assume, a mass of scribbles on a canvas but is inspired by an artistic attempt to eliminate unnecessary elements while retaining the universal. In art theory, abstraction is taught to aspiring artists as the process of stripping down to the essential. It is thought that some abstraction occurs in every form of art, and it is this essence that the artist is intending to elucidate with their work. This is in effect a reversal of nature impinging on our internal model; in art we have an external model reflecting nature's essence. In both domains, we find the first seeds of scientific abstraction, that is, the beginning of detailed perspectives in the development of better modeling systems.

Science is the way we try to decipher our world into manageable chunks. The formalization of abstract thought into symbolic chunks called *language* paved the way for a deeper analysis of reality. We have heard that Aristotle believed that science could not fully interpret reality, that Ockham believed that there were ideas that could not be divined through rational thought, but despite this, the environment influences our model and artists have tried to distill the fundamental in their works. Likewise, scientists have continued to push the bevel incrementally forward. From a strictly scientific perspective, abstractions are the subdivision of data packets, information, into their smallest unit of representation. By systematically categorizing, science has devised abstractions of our world that are superior to the abstractions of the past. Scientific abstraction is meant to better predict the outcomes of our interaction with reality; again, this is achieved by narrowing down to what is essential or universal in a process. One way to do this was to attempt to remove ourselves from the equation, to form a detachment from the principles with which we wish to form theories. The scientific method was one way to achieve that, and this has been done at least twice more in human history with great success. The first was with human language, described earlier in this entry, and the second was with mathematics. Both of these symbolic systems, which distill objects and concepts into universal icons, are fundamentally systems of abstraction.

Mathematics

Though some have argued that mathematics is a self-contained system, no system that is used to represent another system can be defined as anything other than an abstraction. Mathematics is the second most pivotal example of abstract modeling devised by man and has allowed us to visualize, though symbol, scales of magnitude far greater and far smaller than our bodies could comprehend otherwise. From it, we have developed rocket propulsion, physics, and computer science. While arguably more or less important than language in understanding reality, math most likely could not have been derived without it. Indeed, mathematics is perhaps a better abstract modeling tool than language because it is more detached from human emotions and anthropomorphic properties.

Mathematical abstraction requires utilization of everyday words, but in a very precisely defined relationship to mathematically unique symbols. Additionally, mathematics consists of sets of rules for operating on mathematical objects. Therefore, symbolic objects and concepts are manipulated by a specific set of rules to form a self-contained abstract modeling system. Like all abstract systems, the rules and the subjects are generalized; universals represent other structures. Numbers, for example, were initially mathematical symbols that represented quantity. Later, axiomic systems arrived that were considered independent of quantity, which emphasized the abstract nature of the mathematical objects and suggested that they may exist apart from external reference. Again, if you recall, abstractions can be representations of real-world objects, such as a *dozen*, or they can be representations of things that do not physically exist; mathematics reminds us of the nihilistic view that nothing physically exists.

One interesting mathematical concept, which we will see again in computer science, is the mathematical *map*. A mathematical map is a function used to describe a class by a specific parameter, such as the linear transformation in algebra. As such, a mathematical map is an abstraction, which is defined by mathematical abstractions. There are of course many other cases of recursive abstraction. It is math's real-world utility that makes it a powerful abstraction tool, from calculating a monthly budget without cash in front of you to developing a video game using a simulated physics engine.

Psychology

Schema-abstraction, defined in the psychology literature, is the process of learning concepts from examples. Schema-abstraction theories assume that some information is abstracted during learning and stored for classification of new examples. And these theories are distinguished from one another based on how information is characterized, retained, and utilized in the classification of new incidences. In schema-abstraction studies, either a *prototype* model, in which an abstract prototype is constructed and all other examples are compared to it, or an *instance-only* model, in which no abstraction is performed and only training examples are stored, is postulated. Determining the difference between these types of interpretive models has been conducted within psychology research in collaboration with computer science, where the research tools are developed. For example, using a computer mapping program, a series of experiments meant to examine what effect sequence presentation has on conceptual learning was conducted, whereby subjects are asked to devise a rule based on observation of a series of figures given in a specific sequence. However, an explicit definition of the abstraction process has made these studies difficult. For instance, some have argued that learning is dependent on the sequence of presentation of events; others, however, have only seen weak results in experiments looking to replicate this principle. The primary function of schema-abstraction theory investigation is to attempt to delineate which aspects of a model are generalized and when this generalization takes place.

Computer Science

The use of computer programming to generate a number of plausible mappings (*gmaps*) of symbolic object associations has been used in the study of the cognitive abstraction processes described above. Mapping, in computer science, is derived from the mathematics and is used to describe the process of converting associative arrays of data from one form to another. Typically, these programs consist of several modules including the map generator and the structural evaluator, which acts to providing alternate gmap evaluation scores as well as other descriptive modeling tools. Studies in this area have loosely defined abstraction strategies into two classes *limiting case* models, including exemplars only, and *combination* models, including an IF-THEN-Rule Generator. In computer-aided modeling, investigators can investigate several models at the same time. Each model predicts a different type of mental rule formation and can therefore help to define what part of an observation is being processed as an abstraction. Developing investigation tools is only part of interest for computer scientists in abstraction modeling. The binary code of 1s and 0s represents an attempt to distill the information found in analog data into a simpler form. Combining mathematical rules and computer science has led to the creation of intricate simulated environments, sometimes called *engines*, which have been used to analyze everything from engineering to biology. Computer scientists have used abstractions from web browsing histories to predict customer activity, which has been used by search engine companies to target products to consumers. Another lucrative area of research in abstraction is in the field of artificial intelligence (AI). Indeed, understanding the pieces of information used in conceptual abstraction can influence the design of computer systems to create more efficient information processing, one of many goals in creating advanced AI.

From Cognition to AI

Simpler models act as the starting point for more complex modeling. But in contrast to what one might expect, the derivative symbolism from the simpler models allows us to formulate complex models in simpler terms. Like the mathematical map, this is the nature of abstraction. One abstraction can become embedded in another abstraction, reducing the size of the data. By ignoring the irrelevant details, abstract models take many subjects and combine them into one subject. We have seen this in philosophy, where Ockham suggested that the simplest solution usually fits best and called that solution *God*. We have seen this in language, whereby one word, such as *chair*, defines a whole category of objects. We have seen it in math, where one number, for example 5, is the defining factor that generalizes a multitude of instances. We have seen how it can be applied to schema-abstraction models in which either a prototype or exemplar exists. We know that computer scientists are using abstractions today to track the behavior of individuals and market products to each individual accordingly. But these are just the beginning. Abstract models beget more abstract models, which again work to condense multiple objects into symbols representative of them all. The biological process of cognition is thought to use abstraction, which is why it created language and math, and is rooted in the symbolic DNA code of four base pairs, which represent 20 amino acids, from which countless protein structures are generated. Several structures of the human brain are called *pattern generators*. Humans intuitively try to associate objects and concepts into clusters called *patterns*; in other words, we naturally abstract. Perhaps now we can see the purpose of abstraction, a storable code that represents a much more complex system. Data are easier to manage in this simpler form and take up less space. Indeed, humans have extrapolated that the ultimate abstractions would be the ones that define all subjects as one essential category: the unified theory, the Big Bang, God, and so forth. Finally, at the forefront of science, individuals are piecing together the processes found in numerous fields to generate an abstraction of human thought that can fuel an artificial mind. Whereas human cognition is still hundreds of orders of magnitude more efficient than the currently most advanced computer operating system, the growing field of AI has an overwhelming need to define abstractions to fit more information into a finite system. These are the yet-to-be-defined abstract information processing algorithms.

Conclusion

Abstraction, as explained in the preceding sections of this entry, is the process of condensing relevant information and disregarding irrelevant information. It is used in commonsense reasoning in order to simplify conceptualizations and to focus attention on specific details. Our processing of reality occurs in the form of abstraction. Furthermore, language and mathematics, two formalized abstract models, have led in turn to refinement of our abstract characterizations and were required for advanced scientific discoveries in biology, psychology, and computer science. To abstract is to focus on establishing the essential in every system. This is the nature of abstraction: to ever refine and reduce subjects to their simplest and most essential characterizations. Abstraction therefore is one of the most useful tools in theoretical analysis.

Ian C. Clift

See also Artificial Intelligence; Information Theory; Mental Models; Metaphysics; Data Models

Further Readings

Feyerabend, P. (1999). *Conquest of abundance: A tale of abstraction versus the richness of being*. University of Chicago Press.

Fine, K. (2002). *The limits of abstraction*. Oxford University Press.

Loux, M. J. (Ed.). (1979). *The possible and the actual: Readings in the metaphysics of modality*. Cornell University Press.

Wielinga, B., Boose, J., Gaines, B., Schreiber, G., & Van Someren, M. (Eds.). (1990). *Current trends in knowledge acquisition: Volume 8 of Frontiers in artificial intelligence and applications*. IOS Press.

Zucker, J.-D. (2003). A grounded theory of abstraction in artificial intelligence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1435), 1293–1309. https://doi.org/10.1098/rstb.2003.1308

Accuracy

The concept of accuracy is central to the scientific method; this article focuses on how it is employed in scientific practice. In common parlance, a statement is said to be accurate when it is factual. In turn, a statement is factual when it is expressing a proposition that is in agreement with facts or actual states of affairs. In that sense, asserting that a statement is accurate is equivalent to asserting that the statement is true. Moreover, the term *accurate* is often used as a quasi-synonym for other terms such as exact, precise, correct, literal, and similar expressions in the same semantic field. However, in scientific parlance, there are significant differences between the meanings of those terms.

The concept of accuracy is closest to what philosophers of science call *approximate truth*. (Other terms used in philosophy of science with similar meanings include truthlikeness and verisimilarity.) This is reflected in the normative terminological proposals of the International Organization for Standardization (ISO, 2012) where the term *trueness* is used to describe the accuracy of a measurement method (). (The ISO standards also uses the term *precision* to describe accuracy. However, we will not follow this usage here since it conflicts with standard usage in mathematics and computer science, where *accuracy* and *precision* are sharply distinguished.) As such, accuracy is a concept that is related to other methodological concepts such as precision, error, and uncertainty, and this article will sketch how the relation is fleshed out in experimental and computational contexts, which are the two main scientific contexts in which assessments of accuracy are required.

Accuracy is a concept used to characterize statements *semantically*, just like approximate truth. Thus, whereas truth (and accuracy in its colloquial sense) is used to provide a *binary* semantic evaluation—true or false, with no admissible intermediate—accuracy and approximate truth allow for intermediary semantic assessments that admit of degrees. But due to its intrinsic vagueness and lack of general criteria of application, the concept of approximate truth has received some harsh criticisms in the philosophy of science literature. For instance, Laudan (1981) has argued that discussions relying on approximate truth are "just so much mumbo-jumbo." Indeed, what would it mean for the statement *Orgone exists* to be approximately true? Such qualms, however, are dissipated once we observe that approximate truth and accuracy are concepts that do not apply to the semantic evaluation of all kinds of statements.

The scope of application of the concept of accuracy can be rigorously delineated using Carnap's distinction between classificatory (also known as *qualitative*), comparative, and quantitative concepts. Firstly, a classificatory concept is a concept that places an object within a certain class. For instance, when we assert that whales are mammals or that electrons are fermions, or when we make an existentially quantified assertion such as "*dark matter exists*," we are making simple claims about classificatory concepts. Such concepts can typically only be meaningfully evaluated semantically using a binary scale—true or false. Thus, neither the concept of accuracy nor that of approximate truth apply. Secondly, a comparative concept is one for which it is meaningful to provide an ordinal ranking of the degree to which the concept applies. For instance, if we consider the classificatory concept *tall*, we could only assert that a given person is tall or not. With the comparative concept *taller than*, we are now in a position to say that individual *a* is taller than *b*, that *b* is taller than *c*, and so on. However, quantitative concepts are purely ordinal. Thus, the concept *taller than* does not offer resources to assess how much taller *a* is, *b* is, and so on.

Finally, quantitative concepts are those for which there is a numerical value on some scale that is used to determine whether the concept applies. If we consider the concept of distance, we would mathematically specify with a number or a function what the distance of a line segment is. There are many subtleties and complications involved with the selection of a proper scale to assign numerical values as Scott and Suppes make clear in their 1958 essay "Foundational aspects of theories of measurement". However, the point that is important for our discussion is that if we assert

$$\text{The line segment has length } M \tag{1}$$

when in fact it would be true to say that

$$\text{The line segment has length } L, \tag{2}$$

we can give a semantic assessment of statement (1) more refined that just observing that it is false. To the extent that M is close to L, we will say that statement (1) is more or less accurate. We then define the quantity Δ

$$\Delta = |M - L| \tag{3}$$

as the *error*. The smaller the error, the more accurate a statement is. Thus, statements containing quantitative concepts are within the scope of application of the semantic concepts of accuracy and approximate truth.

However, there is a difficulty related to the assessment of the accuracy of a statement that fundamentally entangles it with a related epistemological problem. Indeed, if we knew what the exact (or true) value of the quantitative parameter in the first place, we would have asserted the true statement. Thus, the problem is that we need to assess the accuracy of statements without knowing which value is the exact one. This epistemological problem is central to two methodologically important branches of science, namely, scientific computing and experimental measurement. We discuss them in the remaining two sections of this article.

Accuracy of Computation

The concept of accuracy plays an essential role in computational mathematics because exact solutions to the mathematical equations arising from the construction of realistic models typically do not afford exact solutions. Numerical algorithms are then used to extract information about what the solution to model equations might look like. However, since the computed solutions are typically not exact, an assessment of accuracy is required to validate results.

Computational errors are essentially of three types. Truncation error amounts to replacing functions $f(x)$ (often characterizing vector fields) and integrals $\int f(x)\,dx$ (often characterizing the motion of a body in phase space) by truncated asymptotic series in a perturbation parameter ε, that is,

$$f(x,\varepsilon) = \sum_{k=0}^{N} f_k(x)\phi_k(\varepsilon),$$

for some collection of gauge functions $\{\phi_i\}_{0\le i\le N}$ (see the entry on Perturbation Theory in this encyclopedia). Expressions of this sort have to be truncated, since we often have no closed-form solutions, and it is impossible to add an infinite number of terms in series. Secondly, discretization error is the error incurred by replacing a continuous parametrized flow $\dot{x} = f\left(t, x(t); \mu\right)$ in phase space by a discrete map of the form

$$x_{k+1} = \Phi(t_k, x_k, \ldots, x_0, h, f, \mu).$$

This substitution is the basis for most methods of numerical differentiation and integration. Finally, we typically do not compute the value of functions using field arithmetic (e.g., the familiar arithmetic of real numbers), since computers cannot handle such infinitary entities. Thus, it is replaced with a finite computer arithmetic known as *floating-point arithmetic* (see Figure 1).

In essence, it involves replacing the real line by a *floating-point number line*. As we see in Figure 1(b), it is not really a line; this is why it is important to consider the role of roundoff error in mathematical representation. All of these computational approximations are made because we can only execute finite, discrete operations on a digital computer.

There are three ways of measuring computational error. (For a more extensive explication of these notions, see Corless and Fillion (2013), Higham (2002), or Deuflhard and Hohmann (2003).) Mathematical problems can be thought of as maps from a set of input data to a set of

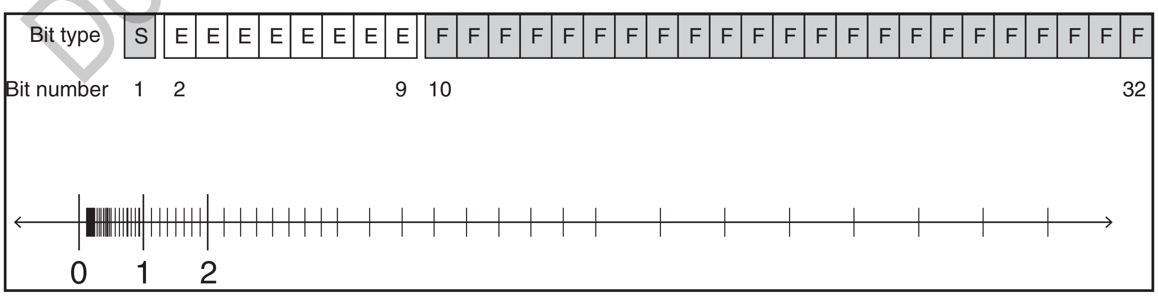

Figure 1 Floating-Point Numbers. (a) Structure of Single-Precision Floating-Point Numbers. There Is a Bit for the Sign, 8 Bits for the Exponent, and the Remaining 23 Bits Are for the Fractional Part. (b) Discreteness of the Floating-Point *Number Line*, With Variable Density

output data, that is, as $\varphi : \mathcal{I} \rightarrow \mathcal{O}$. The mapping itself will typically be given as

$$x \xrightarrow{\varphi} \{y \mid \phi(x, y) = 0\}, \quad (4)$$

where x is the input data, y is the solution sought, φ is the defining function, and φ(x, y) = 0 is the defining equation (or condition). However, as we have seen, when we use a computer for computations, we engineer a modified problem $\hat{\varphi}$ so that it can be implemented and efficiently executed on a digital computer using schemata for discretization, truncated, and rounding off. In general, for a problem such as the one in equation (4), the difference between the exact solution y and the approximate solution $\hat{y}$ will be denoted Δy and will be called the *forward error*. The concept of forward error is typically the one that is referred to when people talk about error; when they refer to a computed solution as being *accurate* or *approximately true*, one means that the forward error is small.

Even if the forward error is the most common, in many contexts this is by no means the most insightful, because we have no direct way of calculating Δ when we don't know what y is. For this reason, there are two other measures of error known as the *backward error* and the *residual*. (For the former, see Fillion and Corless (2014), and for the latter, see the article Perturbation Theory in this encyclopedia.) However, whatever measure of error is used, assessment of accuracy is typically done by reference to the forward error. Since it is typically impossible to obtain an exact quantification of the error, numerical analysts typically analyze the algorithms used and prove that the error must be smaller than a certain quantity or function, known as an *error bound*. As a result, even if the exact error is not known, we know that it lies within a certain interval. If the interval is sufficiently restrictive for the intended purposes, then the computed solution is judged sufficiently accurate.

Accuracy of Measurements

The construction of a model susceptible of generating prediction (or retrodiction) and explanation of the behavior of a system requires the specification of the value for some parameters. Those parameters may be the primitive quantities in which the states of the system are expressed—examples include position, momentum, temperature, pressure, spin, and so on—or some other derived quantity. In either case, the value of this parameter will be supplied on the basis of *measurements*. Nonetheless, no scientist would argue with the fact that all experimental data have some degree of imperfection, in that experimental results always contain errors. As a result, numerical values gathered in experiments are always likely to be wrong (i.e., inexact). This, however, does not imply that the values reported are bad, for they may convey entirely satisfactory information if the measurements are sufficiently accurate. To ensure that inexact values reported are informative, scientists have to diagnose the possible sources of measurement error and must try to design an experimental setup that will ensure that the error is minimized (or satisfactorily small, given what is already known). That is, one must establish the value of a parameter with a methodology that will also make it possible to assess the accuracy of the value.

However, here again it is not possible to exactly quantify the measurement error directly for the value of the parameter being measured is unknown. The key to assessing the accuracy is not to directly quantify the error but to specify an interval within which the error is. Thus, the role of measurements of parameters is to determine (1) a value of the parameter and (2) an estimate of the uncertainty associated with the measurement. The central concepts involved in a theory of measurement are thus the concepts of accuracy, error, and uncertainty.

The discussion to follow is based on the so-called *GUM approach* to the theory of measurement ("GUM" stands for "Guide to the expression of Uncertainty in Measurement"). The relevant technical documentation includes the guidelines of the National Institute for Standards and Technology and the technical reports from the Bureau International des Poids et Mesures. In what follows, the terminology is used according to the International Organization for Standardization (2004).

A measurement is a process involving a *system* and an *apparatus*. The quantity that is subject to measurement is called the *measurand*. The value resulting from the measurement using a certain apparatus is known as the *indication value*. The difference between the value of the measurand (what is often called the *true value*) and the indication value is the *error*. Again, error is a *semantic* notion, relating to a *matter of fact* relating two numerical values. It is not about what we know, ignore, wish to know, or even can know. It must be distinguished from the *epistemological* notion of uncertainty, since "the result of a measurement after correction can unknowably be very close to the unknown value of the measurand, and thus have negligible error, even though it might have a large uncertainty." It is important to stress the distinction since they may *appear* identical as both are understood as intervals within which the *true value* lies. All the scientist can do is to provide an *estimate* of the error based on what is known about the system and the measurement apparatus. ("In general, the error of measurement is unknown because the value of the measurand is unknown. However, the uncertainty of the result of a measurement may be evaluated.)"

This is why, in modern expositions of the theory of measurement, the role of uncertainty is given priority over that of error for the formulation of methodological rules. The first step toward a correct estimation of the uncertainty of the results of a measurement is a diagnosis of the possible sources of measurement error, which are of two kinds: (1) random error and (2) systematic error.

They are also often referred to as *Type A* and *Type B error*, respectively. On the one hand, *random errors* are unpredictable. They are variations in the measurements that the experimenter cannot control (or can control only very limitedly). In terms of probability, it is an error that is just as likely to be above or below the real value; so, for random error, averaging a large number of measured values should, in principle, largely reduce the magnitude of the error. On the other hand, *systematic error* cannot be controlled as random error, that is, averaging will be of no help. Systematic error is caused by the design of the experiment. Its impact can only be alleviated by modifying the design; however, it is often very hard to find a setup that has no systematic error. See Table 1 for examples. Systematic error is particularly problematic, from an epistemological point of view, because there is no way to determine whether there is a systematic error. ("Like the value of the measurand, systematic error and its causes cannot be completely known.)"

Table 1 Examples of Random and Systematic Error

Random Error	*Systematic Error*
Vibration in the floor → fluctuation in balance	During the time required to measure the mass of a fluid, some evaporates
Air currents → fluctuation in balance	During the time required to measure length, the temperature is not controlled and changes
Electrical noise in a multimeter	Miscalibrated balance will cause all the measured masses to be wrong

Each of the uncertainty components that contribute to the uncertainty of the measurement are represented by an estimated standard deviation, termed *standard uncertainty*. This standard deviation may be or may not be evaluated statistically, as Case A or B may be. We talk of *Type A* and *Type B* evaluation of uncertainty. Type A uses any appropriate statistical method. However, the procedure is not as straightforward for the assessment of Type B uncertainty components:

A Type B evaluation of standard uncertainty is usually based on scientific judgment using all the relevant information available, which may include

- previous measurement data,
- experience with, or general knowledge of, the behavior and property of relevant materials and instruments,
- manufacturer's specifications,
- data provided in calibration and other reports, and
- uncertainties assigned to reference data taken from handbooks.

Repeating the experiment can be used to successfully control error arising from random effects. By carefully considering the factors mentioned, *replicating* the experiment can be used to successfully control error arising from systematic effects. (It is important to distinguish repeatability and replicability. Repeatability consists in doing the measurement multiple times with the same apparatus; replicability involves changing the apparatus, each having their own systematic biases.)

The task that remains is to find the *combined standard uncertainty* of a measurement result. The combined standard uncertainty, as its name suggests, combines together all the uncertainty components of Type A and B to generate a total estimate of uncertainty. The usual method used to combine uncertainty is the common statistical method used to combine standard deviations known as *law of propagation of uncertainty*.

We have seen that the concepts of error and uncertainty are intrinsic aspects of the values of the experimentally measured parameters. What does the presence of error and uncertainty imply for the mathematics used to analyze such data? The key challenge here is to develop mathematical techniques that permit us to deal adequately with operations on quantities not known exactly. The most basic approach to do this is that based on *significant figures* (or, alternatively, *significant digits*). The idea is easily understood from the following joke:

> Bazooka Joe is showing a friend a fossilized bone. The friend asks how old it is and Bazooka Joe responds that it is one hundred million and three years old. "How do you know that?" asks the friend. Bazooka Joe responds "The museum expert told me it was a hundred million years old and that was three years ago." (reported by Ruekberg, 1994)
>
> The author then explains the pedagogical relevance of the joke as follows:
>
> The author readily admits that the joke is not *very* funny. That it is funny at all is because even children, just old enough to chew gum without swallowing it, realize that something is wrong about Bazooka Joe's computation: the accurate three years cannot be added to the ball park figure of a hundred million years. (Ruekberg, 1994)
>
> The computation is wrong even if, arithmetically speaking, it is irreproachable. Thus, when there is uncertainty, there is a methodologically important sense of "correctly using mathematics" that differs from the standard one. Understood as a tool occupying a central place in a strategy for the management of uncertainty, the first role of significant figures is to faithfully report the uncertainty in experimental measurements. More precisely, significant figures are a tool to faithfully report the *accuracy* and the *precision* of the results of a measurement, given the *resolution* of the measuring device. It is important to keep in mind the distinction between accuracy and precision. Accuracy is about having the answer right, that is, about having a small error. Precision is about having many digits. For instance, 3.166666666666667 is a very precise (16 digits) but (for many purposes) very inaccurate (two digits) approximation to π. On the other hand, 4.54×10^{-5} is a not very precise (three digits) but quite accurate (precise to the order 10^{-6}) approximation to e^{-10}. Similarly, a measuring instrument can be very precise, and yet inaccurate. As W. Kahan and Joseph D. Darcy (1998) explain, "[p]recision is to accuracy as intent is to accomplishment." A basic objective of the use of significant figures is to not be fooled by measurements that are more precise than accurate; thus, a basic rule is to not report results of measurements with more digits than are accurate, for those extra digits would not be significant. Precision is a property of a linguistic object (namely, of the numeral representing a number in a given number system), whereas accuracy is a semantic notion. Limiting the number of figures reported to the significant figures is a way to make the semantics transparent by showing it in the form of the linguistic expressions used to report results.

The resolution of an instrument is the maximum error that the instrument produces under prespecified circumstances (e.g., value range, ambient temperature, humidity, pressure). (The manufacturers of equipment often indicate how accurately and precisely it can measure.) However, the resolution cannot be smaller than the precision of the instrument. To illustrate this point with a simple example, if a ruler's smallest division is 1 millimeter, then we cannot specify what a length is by measuring with this ruler is to less than half of a millimeter. What we obtain from an instrument with this precision is a number having the format *x.yyz* centimeter, where *x* is the integer part, *yy* are the certain digits, and *z* is the uncertain digit (there is only one of those, the last one). The last digit is only an estimation. Moreover, under the prespecified conditions mentioned earlier, if the instrument is properly calibrated, then each digit within the precision of the instrument is taken to be significant. Accordingly, *the significant figures of a number are those digits that carry meaning contributing to its precision*, and indirectly contributing to its accuracy, provided that some assumptions about calibration are satisfied. (Of course, if the prespecified conditions of calibration are not met, then the reasoning does not hold.) Thus, the significant figures of a number are the digits necessary to specify our knowledge of that number's precision; and in nice contexts, this also reveals the accuracy.

Whereas this first role of significant figures is to *represent* uncertainty numerically by imposing conditions on precision, the second role of significant figures is to permit the formulation of computation rules to determine how uncertainty *propagates*. Those rules are typically formulated in terms of limits on the numbers of significant digits (or significant decimal places) that may be retained in order to faithfully track uncertainty. Here are some standard rules:

1. Exact numbers do not affect the number of significant figures.
2. For addition and subtraction, the answer contains the *same number of decimal places* as the least precise operand used in the calculation. The idea is that you cannot add to or subtract from something not known.
3. For multiplication and addition, the answer contains the *same number of significant figures* as the least precise operand used in the calculation.
4. For logarithms, only those numbers to the right of the decimal place of the operand count as significant.

Such a set of rules constitute what is called a *significance arithmetic*.

Now, these rules should not be thought of as being perfectly reliable. Rather, they work as rules of thumb. One might be surprised that there are, well into the 20th century, many publications debating how significant digits should be analyzed and understood. This is because for any set of such rules based on significant digits it would be relatively straightforward to generate problematic cases. Thus, we see that as an attempt to provide context-independent syntactic rules meant to support management of uncertainty and its propagation, significance arithmetic has limitations. More refined mathematical methods of assessing error and uncertainty propagation in order to properly assess accuracy in a more robust way include sensitivity analysis and perturbation theory.

Nicolas Fillion

See also Data Models; Measurement; Perturbation Theory

Further Readings

Carnap, R. (1966). *Philosophical foundations of physics: An introduction to the philosophy of science* (300 pp.). Basic Books.

Corless, R. M., & Fillion, N. (2013). *A graduate introduction to numerical methods, from the viewpoint of backward error analysis* (868 pp.). Springer.

Deuflhard, P., & Hohmann, A. (2003). *Numerical analysis in modern scientific computing: an introduction* (Vol. 43). Springer Verlag.

Fillion, N., & Corless, R. M. (2014). On the epistemological analysis of modeling and computational error in the mathematical sciences. *Synthese, 191*, 1451–1467.

Higham, N. J. (2002). *Accuracy and stability of numerical algorithms* (2nd ed.). SIAM.

International Organization for Standardization. (2004). *International vocabulary of basic and general terms in metrology (VIM)*. Technical Report DGUIDE 99999. Revised from the original 1993 edition.

International Organization for Standardization. (2012). *Accuracy (trueness and precision) of measurement methods and results, Part 1*. Technical Report 5725-1. Revised from the original 1994 edition.

Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data—Guide to the*

expression of uncertainty in measurement. Technical Report JCGM 100:2008. Bureau International des Poids et Mesures. Revised Edition of GUM 1995.
Joint Committee for Guides in Metrology. (2009). *Evaluation of measurement data—an introduction to "guide to the expression of uncertainty in measurement" and related documents*. Technical Report JCGM 104:2009. Bureau International des Poids et Mesures.
Kahan, W., & Darcy, J. D. (1998). *How Java's floating-point hurts everyone everywhere* [Presentation]. ACM workshop on Java for high-performance network computing, Stanford University, CA.
Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science, 48*, 19–48.
Ruekberg, B. (1994). A joke based on significant figures. *Journal of Chemical Education, 71*(4), 306.
Schwartz, L. (1985). Propagation of significant figures. *Journal of Chemical Education, 62*(8), 693.
Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic, 23*, 113–128.
Taylor, B. N., & Kuyatt, C. E. (1994). *Guidelines for evaluating and expressing the uncertainty of nist measurement results*. Technical Report NIST Technical Note 1297, National Institute of Standards and Technology.

Actor–Network Theory

Actor–network theory (ANT) developed within the wider interdisciplinary field of science and technology studies (STS), and more specifically, out of debates within the sociology of scientific knowledge (SSK). SSK, in a series of case studies of scientific controversies, demonstrated that the acquisition of scientific knowledge could not be fully explained by its logic, method, or normative institutions. Taking the work of Thomas Kuhn (1922–1996) as a starting point, SSK demonstrated that the very content of scientific knowledge, including mathematics, was socially constructed. As Bruno Latour (1947–) most explicitly argued, however, SSK inadvertently replaced the natural determinism of the philosophy of science with a social determinism that was no more tenable. In addressing this criticism, ANT's seminal case studies such as Latour's *The Pasteurization of France* and Michel Callon's "Some Elements of a Sociology of Translation" followed the process of *heterogeneous engineering* by which technoscientific objects and truths were assembled and made durable through networks of human and nonhuman actors. Thus, ANT extended SSK's symmetry principle, which held that the same types of causes should be used to explain both true and false beliefs, to a *generalized symmetry* that treated humans and nonhumans as actors with agency. Though the exact construal of nonhuman agency within ANT remains subject to debate, largely around whether to preserve intentionality as an analytically distinct feature of human action, ANT immediately gained attention as an approach to studying how *the social* is continuously enacted as relational networks of humans and nonhumans. In ANT's subsequent articulations as a general framework for studying social order, three of its concepts, *heterogeneity*, *material relationality*, and *punctualization*, initiate a radical reconceptualization of agency, social action, and society itself. This entry addresses each concept, in turn, as well as its consequences for humanistic social theory.

ANT insists that social phenomena are *heterogeneously* constituted. That is, its objects of inquiry—whether scientific facts, nation-states, organizations, and so on—are made up of both human and nonhuman components. ANT emphasizes that a potentially endless array of *actants*, a term borrowed from literary theory and semiotics to designate any human or nonhuman that can accomplish action in a network, may constitute an object of interest. An aspect of ANT's empirical concern thus lies in identifying, or *following*, the specific actants that are relevant in shaping the phenomena in question. In the case of California state formation, for example, Patrick Carroll argues that it is one of the most productive agricultural centers in the world only so long as levees, dams, irrigation, farmers, metering devices, state agents, and so on, are networked together. Heterogeneity greatly distinguishes ANT from humanist social sciences that, in this example, would overlook the relationship between state territories and their material infrastructures; that is, how the historicity of material culture shapes the state of California. Under an actor–network formulation, then, the state's ontology is composed of heterogeneous elements rather than viewed as a single, coherent actor that is distinct from society. So, an initial premise of ANT is that *the social* is made up of a heterogeneous collection of humans and nonhumans, though their relationality is always an empirical question.

To the heterogeneity of the social, add ANT's concern for *material relationality* and this is how ANT proposes to study the social world: not from *a priori*, binary divisions between humans/nonhumans, nature/culture, state/market, and so on, but in terms of how such phenomena are constituted in the continuous unfolding of heterogeneous relations. This second actor–network concept argues that entities have no essential reality apart from the web of relations in which they are produced. In other words, social phenomena are generated by human and nonhuman actor–networks and it is the relationality between heterogeneous elements that ANT's case studies pursue. This concern for relationality orients ANT toward the *way things are*, the ontological character of the heterogeneous actor–networks through which the world is made. ANT argues that what gives objects, forms of knowledge, and other social phenomena their meaning, stability, or power, must be examined in terms of how their *bits and pieces* are held together.

In the decades-old project of damming the Mississippi River, for example, an effective levee-building science develops in relation to the resistance of the river itself. Andrew Pickering captures this relational process as a *dance of agency* between engineers who build dams, a river that resists their control, and the subsequent redesign of their materials. The result is an engineering science that has become increasingly more attuned to the sedimentary composition and currents of the river. ANT case studies emphasize this always contingent, never final, way in which humans and nonhumans form more durable wholes through complex material relations. In this sense, ANT shares an emphasis on the unfolding and processual dimensions of the social world characteristic of poststructural, ethnomethodological, and pragmatist philosophies.

When taken together, heterogeneity and material relationality are significant both for the empirical study of scientific practice as well as for ANT's theory of social action. With respect to the former, ANT suggests that the work of technoscientists is not simply to follow a method, conduct experiments, and reveal the *truth* about the natural world. Rather, technoscience is a complex and strategic practice in which the diverse materials encountered must be *assembled* into a durable, material network that can overcome unforeseen obstacles. For example, Bruno Latour's seminal case study rejected the familiar story that attributes Louis Pasteur's success to his genius mind. Instead, Latour pointed to Pasteur's technoscientific practice, the material arrangement of a heterogeneous actor–network composed of laboratories, domesticated bacteria, statistics, notebooks, microscopes, and more, as the explanation for his triumph. Pasteur's success, in other words, is an effect of organizing these heterogeneous components into a stable, material arrangement. Thus, in ANT's reformulation of social action, it was not Pasteur's rational mind that distinguished him from other technoscientists, but the collective strength of the heterogeneous actants he *assembled* as a network. This insight further supports the STS thesis that science is a social and cultural activity, and in a final move, further reconceptualizes social action.

The third concept, *punctualization*, refers to the process by which a heterogeneous network stands as a singular and coherent object or actor. It suggests that when *Pasteur-the-great-researcher* is evoked as a scientific hero it is as an effect of his heterogeneous network. Crucially, punctualization entails an oscillation between powerful individuals or objects and the contingent, network-building practices through which they become recognizable as such. The concept can thus be viewed as an alternative to human-centered histories and philosophies of science that tend to *black box* the process of knowledge production; that is, attribute knowledge itself to the brilliance of a solitary individual. As ANT instead directs empirical attention to precisely the activities through which a matter of fact is successfully constructed, it explains Pasteur's success as a heterogeneous and collective achievement, an effect of the associations of multiple heterogeneous actants. Under an actor–network formulation, Pasteur cannot become a brilliant technoscientist of eventual world renown apart from a durable network. Still more radically, Pasteur's achievement *is* the collective achievement of a heterogeneous actor–network for which he, when evoked as a singular individual, becomes a *spokesperson*.

In conclusion, ANT argues that agency, social action, and society itself are empirically complex and in need of explanation. Whereas humanist social sciences tend to privilege a social world made up of intentional, human actors and defined by boundaries such as state/society, nature/culture, and social/scientific, ANT proposes an ontology in which nonhumans and humans are analytically symmetrical in collective assemblages. It is these collective assemblages, in their complex and material relationality, that ought to be explored. ANT's concepts of heterogeneity, material relationality, and punctualization, lead to a *distributed* conception of social action concerned with a multiplicity of human and nonhuman actants. Such a thesis not only moves ANT closer to what feminist STS has considered *post-humanism*, it also orients its analytic tools to the ontological complexity of action in the world.

Sam Haraway and Patrick Carroll

See also Pragmatism; Social Construction of Scientific Knowledge

Further Readings

Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and fishermen of St. Brieuc Bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge?* (pp. 196–223). Routledge.

Carroll, P. (2012). Water and technoscientific state formation in California. *Social Studies of Science*, *42*(4), 489–516.

Latour, B. (1988). *The pasteurization of France*. Harvard University Press.

Law, J. (1992). Notes on the theory of the actor network: Ordering, strategy and heterogeneity. *Systems Practice*, *5*(4), 379–393.

Pickering, A. (2008). New ontologies. In A. Pickering & K. Guzik (Eds.), *The mangle in practice: Science, society, and becoming* (pp. 1–15). Duke University Press.

Ad Hoc Hypothesis

The topic of this entry, the ad hoc hypothesis, is a small and recent footnote in a long tradition of debating how best to determine scientific truth. On the surface, the concept is straightforward: a hypothesis is formed *ad hoc*, that is, in response to a particular situation instead of in anticipation of a general application. A host of validity issues are prompted by the introduction of an explanation ad hoc: validity of the theory, of the investigation, of the findings, and of the generalizations that may follow. These issues are the context in which the term must be understood, beginning with the fact that it is usually associated with pseudoscience, that is, conclusions which cannot be verified independently. That said, it is important to recognize a new context in which ad hoc hypotheses figure: belief revision within computer programming for artificial intelligence.

The term itself has prompted great debate concerning its definition, its purpose, and its value. Let us first consider the term *ad hoc*, a Latin phrase meaning *for this*, which is used in reference to something that is *one-off*, that is, not an established and predictable routine, but formed for a specific and short-term purpose. An ad hoc committee functions as a task force with a brief to address one issue, and then it is disbanded once the issue is addressed. The colonial militias were ad hoc because they formed in response to an immediate threat after which the solders reverted to their chosen occupations. In contrast, a standing committee or a standing army will have well-defined continuous roles in relation to the agencies they serve, and their members develop more professional expertise, whereas ad hoc groups operate on the common sense and skills of the average agency member. This adds to the more casual association attached to *ad hoc hypotheses*, which are also considered less rigorously developed and less valuable. An ad hoc hypothesis is produced to account for anomalies that are inconsistent with the original hypothesis.

Hypotheses are propositions of general truth, that is, a prediction of an outcome that will naturally occur based on the logic of a theory or belief. These statements are formal interpretations of the hunches and guesses that prompt researchers to investigate phenomena. Hypotheses are used to limit the scope of the investigation that is intended to test the validity of the theory prompting the research question. If the findings are inconsistent with the predictions of the theory, the theory itself is then in doubt. The value of a theory lies in its capacity to explain naturally occurring phenomena and to help people predict and perhaps control future events. This reasoning is based on a positivist orientation to reality, meaning that the world is understood to be fairly stable and, given enough rigorous observation, its patterns can be recognized with some confidence and in turn used with confidence in similar situations. At issue is the presence of anomalies that are inconsistent with the theorized expectations, and whether they should be used to discard the theory or inform the theory. Empirical investigations are designed to test hypotheses in a systematic way so that others can replicate the procedure and thereby test the validity of the findings. Adding weight to this debate is the pride of integrity that prompts many researchers to distance themselves from practices inconsistent with their epistemological orientation and even to disparage those who appear to compromise the quality of rigor they seek.

It is safe to say that there is a keen competition among academics to claim authority. A theory is considered less robust if it cannot withstand general application without special adjustments to different circumstances. Whether termed ad hoc hypotheses or rival hypotheses or counterarguments, statements are considered fallacies if they are based on factors that cannot be tested, such as paranormal activity, or *fate*, or superstition. They may be dismissed as simply unhelpful because they distract from the task at hand or may be found unattractive because they disregard the elegance of a simple equation. They rarely add persuasive plausibility. As found in the political arena, ad hoc explanations are called *spin control* and are associated with a cynical manipulation of language to camouflage the failure of the findings to be self-evident. Within a peer review of an investigator's report, there will be keen criticism if the discussion introduces theory that was not established in the literature review justifying the research hypothesis. There is instead a convention of intellectual humility: to identify future research trajectories given the limitations of the study.

In the larger context of advancing scientific knowledge, however, the practice can be viewed as developmental in the professional researcher. The rigidity of a novice's insistence on the absolutes of binary opposites gives way to the flexibility of an expert who can tolerate ambiguity, paradox, and multiple explanations in the process of evaluating the merits of a theory. The naïve and inexperienced are motivated to defend a preconceived notion, while the more experienced will consider the pattern of empirical evidence. These are in fact the argumentational and rhetorical skills necessary for scientific discourse: generating evidence and counterarguments to establish proof and certainty. Ad hoc hypotheses are a function not of the original experimental design but of its interpretation by the investigator. The investigator may believe that this modification improves the explanation of the phenomenon, but the more traditional scientific community tends to regard this as an evasion, motivated by a desire to protect the theory in question from disputation.

The role of ad hoc hypotheses could be absorbed into the fabric of basic scientific discourse were it not for a current interest in mechanizing the executive function of decision. *Artificial intelligence* is a contemporary field attempting to program computers with all possible contingencies so that the program can actually *learn*; that is, revise its own procedures based on responses. A computer program is a series of propositions requiring the programmer to predict the nature of decisions to be made, contingent on other actions. The algorithm by which the program proceeds through these decisions is of course crucial to the outcome, for a simple reversal of steps can result in a significant change. *Belief revision* in this context does not mean a change in philosophical conviction about reality or morality, but a change in technological procedures that are, from the standpoint of the machine, the way things are. These are fundamental issues for the organization of information within all fields, so a belief revision theory launched a wide-ranging field of research. This theory is found in the Alchourrón, Gärdenfors and Makinson (AGM) model, a formal framework for analyzing ways to change beliefs in the context of computer programming. The three researchers collaborated from different fields to propose it, illustrating the multidisciplinary interest in this logic. Complex mathematical formulae are the currency of this discourse, akin to setting up an intricate series of dominoes in readiness for a simple action to trigger a long sequence of response. This enterprise assumes a closed system of variables; that is, it operates within a set of propositions, but as databases accumulate, they must be updated, leading to an important concept of database priorities within truth maintenance systems. They are also deterministic in that given a set of beliefs responding to some input, the resulting belief set is well anticipated. A combination of operators may be used as an indeterministic operator that allow more than one admissible outcome. Ad hoc hypotheses are thus a mainstay of the dynamic process.

Ad hoc hypotheses highlight the epistemological distinctions between belief and knowledge, between intuition and theory, and between deductive and inductive approaches to problem-solving. Opinions about what ad hoc hypotheses are, and their value, reflect one's orientation to identifying truth based on observed phenomena, on principles derived from an accepted theory; for example, syllogisms associated with Aristotle, or on authoritative conclusions such as those demonstrated in Socratic dialogues, which used hypothetical situations for support of interpreting ideals. Modern rationalism follows René Descartes' four steps of scientific thinking: skepticism (that nothing is true until proven so), reductionism (divide concepts into much smaller entities that are more easily observed), organization (ordering knowledge from simple to complex), and generalization (refining a theory to eliminate omissions and contradictions). Francis Bacon argued that theory must be built from the ground up by first accumulating a large sample of data, then proposing some preliminary axioms or predictable patterns, and finally generating a theory that would explain all the phenomena.

Naomi Jeffery Petersen

See also Falsifiability; Justification; Refutation; Scientific Realism

Further Readings

Bamford, G. (1999). What is the problem of ad hoc hypotheses? *Science and Education*, *8*(4), 375–386.

Bamford, G. (1993). Popper's explications of 'Ad Hoc'ness: Circularity, empirical content, and scientific practice. *British Journal for the Philosophy of Science*, *44*(2), 335–355.

Hunt, J. C. (2012). On ad hoc hypotheses. *Philosophy of Science*, *79*(1), 1–14.

Klein, E. E., & Herskovitz, P. J. (2007). Philosophy of science underpinnings of prototype validation: Popper vs. Quine. *Information Systems Journal*, *17*(1), 111–132.

Patton, L. (2012). Experiment and theory building. *Synthese*, *184*(3), 235–246.

Analysis

According to the prevailing definition, analysis is a process that involves breaking something up into simpler or component parts. Indeed, the word *analysis* derives from the Greek *ana* for up and *lysis* for loosening or separation. But analysis can also be conceived as the search to determine prior principles, beginning from what is known—a conception with roots in ancient geometry and with purchase in academic philosophy. Analysis can be contrasted with its converse, synthesis, which is the process of joining parts or elements together to create a more unified result. Both synthesis and analysis are directed toward the ends of elucidation; through analysis, that which is being analyzed becomes better or more fully understood.

Analysis is a fundamental mode of inquiry and is applicable without restriction; one can conceivably analyze anything. As a result, it occurs throughout lay and specialized academic contexts. In their daily lives, people regularly engage in analysis, analyzing the texts they read or watch, the events they attend, the words, deeds, and behaviors of other people, and so on, discerning component parts of larger wholes. Similarly, scholars employ analysis across a wide range of academic contexts, including the sciences, social sciences, arts, business fields, and in math and the humanities, where it is a basic building block of intellectual pursuit. Scholarly analysis is generally distinguished from lay analysis by a greater degree of rigor and formalization, though the extent to which it is either depends on the particular discipline and method.

As a decompositional process, analysis includes both descriptive and pattern-seeking activities. As descriptive, analysis involves identifying the parts of a whole and providing an account of the relevant features of those parts, in some greater or lesser degree of detail. As a result, it is closely tied to observation; and the observation that drives analysis can involve any of the senses: visual, auditory, gustatory, tactile, or olfactory. Moreover, in order to facilitate analysis or, in some cases, even make it at all possible, one might utilize instruments to augment one or more of the senses. The analysis of physical entities or qualities, for examples, often involves the use of instruments or tools. But perhaps more often than being purely descriptive, analysis also entails the search for patterns. Breaking a larger and more complex whole into smaller and more elementary parts brings the interactions between those parts into focus. And in analyzing, one classifies and groups, identifying relationships and drawing conclusions about them.

Analysis is a comprehensive term that refers to a highly generalized activity, which can be actualized in innumerable different ways. For example, analysis can proceed at any level of abstraction. While a biologist might probe a tangible organism in order to analyze its concrete parts, philosophers use conceptual analysis to examine thoroughly abstract concepts like *truth* and *knowledge*. Levels of granularity also vary widely; coarse-grained macro-analyses assess relatively large-scale components, while fine-grained micro-analyses consider relatively small ones. One can even analyze analyses, as is the case in a meta-analysis. Indeed, there are innumerable different approaches and/or specific formalized methods one can employ in order to analyze a target of interest, some of which are outlined in the subsequent sections of this entry.

Analysis is an inherently interpretive activity. In the processes of analyzing, one makes decisions about where to direct attention, about what constitute the meaningful parts of a whole, and construes the existence of relationships and patterns. And while such judgments are sometimes relatively clear-cut, they can also be uncertain and arguable. In the act of breaking something down into its component parts, therefore, we explain how to understand it and make a case for what it means. We render it otherwise. As a result, the terms *analysis* and *interpretation* are sometimes used synonymously. For example, an analysis of a poem might also be described as an interpretation of it. Interpretation, however, does not have the decompositional emphasis that analysis does. Moreover, interpretation often tackles questions of meaning that fall outside of the purview of analysis. Similarly, analysis differs from criticism, another related intellectual pursuit, in that it is not in essence evaluative.

An analysis is not only the act of breaking something down into component parts but also the communicated record or result of so doing. The form of that resultant analysis depends on the type of analysis one has done and the type of context in which one has performed it.

Deductive Approaches

In deductive approaches to analysis, an existing hypothesis, theoretical model, or conceptual framework guides the descriptive and pattern-seeking activities of analysis. Therefore, the hypothesis, theory, or framework directs attention, determining to some extent the parts, patterns, and relationships delineated. As such, they can be characterized as offering a perspectival lens. For example, in a hypothetico-deductive experimental paradigm, one analyzes data in relation to a preidentified hypothesis to determine whether or not it supports that hypothesis. Similarly, a deductive analysis of a dream guided by Freudianism will focus on features specified by or emphasized in Sigmund Freud's psychoanalytic theory, such as repressed memories and unconscious desires.

Within disciplines, certain conceptual frameworks tend to predominate, becoming frequently engaged as analytic tools. In the humanities and social sciences, for example, psychoanalysis, Marxism, feminist theory, structuralism, critical theory, phenomenology, and deconstruction have all been commonly employed toward the ends of deductive analysis. Those employing these and other models in a deductive fashion must guard against confirmation bias, or the tendency to only look for/see that which confirms an existing hypothesis, overlooking those features that do not fit within its parameters. Similarly, deductive approaches are vulnerable to what is sometimes called *cooker-cutter criticism*, whereby a theoretical framework so strongly and narrowly governs analysis that it becomes a rote and mechanical application of theory with predetermined and not particularly insightful results.

Although an explicitly acknowledged and articulated theoretical framework drives some analyses, the principles that guide analysis in a deductive fashion are sometimes implicit. This is particularly true in lay contexts, where neither theoretical frameworks nor the process of analysis are typically formalized, but also occurs in scholarly contexts as well. Moreover, folk theories sometimes deductively underlie scholarly analysis in a covert fashion. The mobilization of implicit theoretical frameworks can be problematic given that the premises by which the analysis is guided are not articulated and may not be apparent to the person doing the analysis and/or the person encountering the results of the analysis, making them unavailable for discussion or critique.

Inductive Approaches

In inductive analysis, inquiry begins with the analysandum itself; one works with and from the object of interest in order to discern its meaningful parts and draw conclusions about the relationships between those parts. That is, the *breaking down* of this form of analysis involves a *building up* from particulars as one moves from specific features to more general conclusions about parts and patterns. An example of a commonly employed inductive method of analysis in the humanities is close textual analysis, which involves paying scrupulous attention to the details of a selected work or works, from the use of commas to the overarching organization, to identify patterns within that work and/or among works.

However, though theory does not drive an inductive analysis, it must inevitably be guided by something, which often includes intuition or past experience with other texts. This raises the question of the extent to which any analysis can in fact shelve theory, and whether inductive approaches are in fact inevitably theory-laden, even if that theory is of the informal and implicit kind that comprises intuition and develops out of experience. Moreover, philosophical debates about and psychological investigations into the theory-ladenness of observation are clearly applicable to analysis given its reliance on observation.

The human impulse to see and seek patterns and connections not only results in a general affinity for analysis, but also a propensity to see patterns and perceive connections where there are none. This tendency underlies many superstitious and supernatural beliefs and is responsible for much of what is termed *magical thinking*, but it is also a tendency with implications for scholarly analysis. *Overreading* or *overinterpretation* are terms sometimes used to describe overly aggressive acts of analysis in which patterns are read into a text rather than being found there; and this problematic analytic misstep can be found in all sorts of analyses, from the poetic to the radiographic.

While deduction and induction are distinguishable orientations toward analysis, in practice the two often co-occur. They do so in a compelling way in the *grounded theory method*, a formalized methodology developed in the social sciences by which one self-consciously and systematically puts inductive analysis toward the ends of theory development. In using analysis to arrive at theory, one reverses the typical or traditional pattern by which theory precedes analysis. It is an iterative process in that one repeatedly moves back and forth between data and analysis, and between inductive and deductive modes of thought, becoming increasingly abstract and theoretical as the process unfolds.

Quantitative Analysis

Quantitative analysis is the analysis of that which has been quantified; that is, calculated, counted, or measured. And in practice, it can take many different forms. Perhaps the simplest type of quantitative analysis is the discernment of relative quantities. A quantitative chemical analysis of this type, for instance, might determine the amounts of the various component parts of a more complex chemical. But quantitative analysis often moves beyond the relatively straightforward assessment of amounts, to discern patterns and relationships, which is a way of making numerical data meaningful and/or useful. Quantitative analysis can also be put toward the ends of prediction. In predictive analysis, one looks at what has happened in the past and identifies patterns in order to predict future events, as in analyzing voting behavior in previous election cycles to predict outcomes in upcoming ones.

Statistical methods are mathematical tools for seeking out and delineating the patterns in a quantified subject of analysis. Indeed, statistics, a branch of mathematics, is the science of quantitative analysis. Statistics can be applied in a deductive fashion, as in confirmatory data analysis, whereby one identifies a hypothesis and employs statistical analysis in order to test it. Or one can use statistics in a more inductive fashion, applying them to a set of data in order to see what patterns emerge, a type of exploratory data analysis. There are two main branches of statistics, descriptive and inferential. *Descriptive statistics* are methods for characterizing the features of a given set of data; these include measures of central tendency (i.e., mean, median, and mode) and spread (i.e., range, variance, and deviation). *Inferential statistics* are methods for using a sample in order to identify the parts, patterns, and relationships of a larger population; in doing so, the quality of the sampling process is key. *Regression analysis* also deserves mention as a set of powerful statistical processes for examining the relationships between variables, often independent and dependent ones.

In some quantitative studies, particularly those in the social sciences, coding is an important initial step in the analysis process. It is the procedure by which one applies shorthand labels to data items, such as parts of interview transcripts or field notes; in doing so, one not only categorizes them and enables them to be more readily sorted, but also makes them countable and thereby amenable to analysis via quantitative techniques. For example, content analysis, a technique for analyzing communicative acts, often employs coding toward the ends of quantification. When coding, one must decide on the unit of analysis (e.g., words, paragraphs, parts of images, entire photographs) and whether to use predetermined codes or to derive codes during the course of coding. The process of coding data does some analytical work in that it discerns recurring features. It also facilitates subsequent *higher-level* analysis. The reliability of the coding process is an important consideration, and studies often consider *intra-coder reliability*, or the consistency with which a given person codes the data, and to enhance reliability, some studies employ multiple coders, and consider *inter-coder reliability*, or the extent to which they agree with each other.

Visualization is also an integral part of quantitative analysis. The results of quantitative analysis are often effectively presented in visual form, which can incisively convey patterns and relationships between component parts. Tables and graphs, frequency distributions, bar charts, scatterplots, and histograms are all commonly used to present the results of a quantitative analysis. Moreover, data visualization is itself considered a method of quantitative analysis and a branch of descriptive statistics. As a method of analysis, we can consider the process of visualization itself as that which discerns meaningful components and their relationships.

As a result of advances in computing, quantitative analysis has become not only more large-scale and sophisticated, but also increasingly democratized. While the computing power necessary to do complex statistical analysis used to be in the hands of only a few, it is now much more widely available. Moreover, there is now unprecedented access to vast data sets, such as those generated via new media, offering opportunities large-scale quantitative analysis with considerable potential across academic disciplines and for commercial purposes. In professional fields, the word *analytics* is often used to refer to the science of data analysis.

Qualitative Analysis

Qualitative analysis is the analysis of that which has not been (and perhaps cannot be) measured or quantified. It focuses on qualities and features rather than amounts. For instance, a qualitative chemical analysis might detail the characteristics of the component parts of a more complex compound. As with quantitative analysis, qualitative analysis can proceed inductively or deductively; that is, one can take the detailed particulars of the analysandum as the analytic starting point or can proceed on the basis of a theoretical framework.

While the phrase *qualitative analysis* might well be used to identify a wide range of analytic inquiry, it is generally associated with a research paradigm in social science fields that emerged during the 20th century to develop and systematize qualitative methods as legitimate alternatives to positivistic and quantitative ones. As a result, not all of those who engage in analysis that could certainly be described as qualitative would necessarily use that phrase to describe their work, particularly those who work in the humanities. For example, philosophy's tradition of conceptual analysis, which is solidified in the branch of philosophy known as *analytic philosophy*, is certainly not quantitative, but philosophers would not tend to identify it under the banners of *qualitative analysis* and *qualitative research*.

In the social sciences, qualitative analysis tends to be rather systematized and is often part of a research process with parallels to scientific research. For example, the qualitative data are often systematically generated or collected for the purposes of analysis; this is in contrast to the analysis of something preexisting and simply selected for study, as in a rhetorical analysis of an influential speech or the art historian's analysis of a famous painting. Qualitative researchers in the social sciences might conduct interviews, run focus groups, and engage in participant observation, analyzing the transcripts or notes that result from these activities. These textual data are also often coded, a process that is not only an instrument of quantification but is can be a means of sorting data and thereby facilitating qualitative analysis. Some of the specific approaches to qualitative analysis that are widely used in the social sciences include narrative analysis, discourse analysis, phenomenological analysis, and grounded theory.

Qualitative and quantitative analysis can be complementary ways of discerning the parts and patterns of a subject of interest, and many research projects, such as mixed-methods studies, use both. Moreover, social scientists sometimes use qualitative analysis in order to engage in initial exploration of a subject, followed by quantitative analysis to test hypotheses derived from those exploratory qualitative endeavors.

Critiques of Analysis

Alongside the widespread embrace of analysis as fundamental mode of intellectual engagement exists a strain of skepticism regarding its power to elucidate. The so-called *paradox of analysis*, associated with C. H. Langford and G. E. Moore, is, essentially, the idea that to give a correct analysis of a concept, one already has to know and be able to identify its component parts. For instance, Moore provides the example of analyzing the concept of brother into male and sibling, something that requires already knowing what a brother is. The analysis is trivial and does not provide any new insights. As a result, it follows that it is impossible for an analysis both to be correct and to inform us of something we do not already know. And although debates regarding paradox of analysis (which include attempts to resolve it) occur specifically in relation to philosophical conceptual analysis, the paradox has relevance to analysis more broadly considered.

Others express concern regarding analysis's essential reductiveness; by definition, analysis involves breaking things down, distilling the elementary parts from the more complex whole. Philosophers such as Georg Wilhelm Friedrich Hegel and Friedrich Schiller have pointed toward the reductive dimension of analysis. Pierre Teilhard de Chardin once described analysis as "that marvellous instrument of scientific research to which we owe all our advances but which, breaking down synthesis after synthesis, allows one soul after another to escape, leaving us confronted with a pile of dismantled machinery, and evanescent particles" (1959, p. 283). That is, the elucidation of analytic dismantlement is not without its losses.

Michelle G. Gibbons

See also Abstraction; Conceptual Analysis; Generalization; Interpretation

Further Readings

Beaney, M. (2014). Analysis. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2014 ed.). Retrieved July 2014, from http://plato.stanford.edu/archives/sum2014/entries/analysis

Berelson, B. (1952). *Content analysis in communication research*. Free Press.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

King, J. C. (1998). What is a philosophical analysis. *Philosophical Studies, 90*, 155–179.

Miles, M., Huberman, B., Michael. A., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE.

Selzer, J. (2004). Rhetorical analysis: Understanding how texts persuade readers. In C. Bazerman & P. Prior

(Ed.), *What writing does and how it does it: An introduction to analyzing texts and textual practices* (pp. 279–308). Erlbaum.
Teilhard de Chardin, P. (1959). *The phenomenon of man* (Trans. B. Wall). Harper.
Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Applicability of Mathematics in Science

Our understanding of nature, our explanations of natural phenomena, crucially depend on the use and interpretation of mathematics in scientific practice. Mathematics provides a common language for the empirical (i.e., the natural and social) sciences, and through its tools, we get successful predictions and empirical confirmations that are about the entities and the phenomena of these sciences. When this happens, we say that mathematics successfully applies in science. But the success of mathematics in science prompts the following question: How is it possible for the abstract entities of mathematics (numbers, equations, integrals, etc.) to say something interesting about phenomena and entities, as for instance nuclear reactions and proteins, which are studied by the empirical sciences and have a completely different nature? The issue is philosophical, and the research topic associated to it is known as the *Applicability of Mathematics in Science*. Following a short introduction to the topic, this entry offers an overview of some accounts that have been proposed to tackle the philosophical issue. It also identifies some connections that have been established between the study of the applicability of mathematics in science and other debates in philosophy of science and philosophy of mathematics.

The Philosophical Challenge

When will Mars be closest to Earth? What is the distance covered by a falling body in a given time under the force of gravity? Given a protein with a specific amino acid sequence, what will its three-dimensional structure be like? What is the average number of predators in a particular ecosystem? How can we maximize profit in a market economy with production constraints? Although these questions involve entities (planets, falling bodies, proteins, predators, production constraints) and empirical sciences (physics, computational biology, ecology, economics) that are quite different from each other, there is a common path to answer them: Do the math! Surprisingly, mathematics works. This, of course, does not mean that mathematics always works when it is used, or *applied*, in the empirical sciences. Nevertheless, in many cases, mathematics provides a correct, empirically accurate, answer to our original question about planets, proteins, predators, and so on. Why should we be surprised about such effectiveness, or successful applicability, of mathematics? The issue is philosophical: Mathematical entities are generally regarded as abstract because they do not causally interact with anything (e.g., a number cannot break a glass) and have no spatiotemporal location (e.g., a number cannot be found in a certain place at a certain time); therefore, it is quite surprising that they can properly inform us on the behavior and properties of entities like planets or proteins, which are causally active and are located in space and time. Even more surprisingly, sometimes the mathematics that is successfully used in science is not developed from scratch to study an empirical problem, but it has been developed within a purely mathematical framework before being used with success in a specific context of application (the mathematics of group theory, for instance, was developed long before its introduction into quantum physics in the late 1920s).

It is exactly this particular feature of mathematics, namely that of being applicable with success outside its boundaries, that allows scientists to make extraordinary progress in the understanding of nature. And the main goal of philosophers of science and mathematics working on the applicability of mathematics in science is to investigate such features and account for the (successful) application of mathematics in science.

Facing the Challenge

The applicability of mathematics has a longstanding philosophical pedigree. Indeed, the first reflections on the successful use of mathematics to understand the world can be traced back at least to early Pythagoreanism (sixth through fourth centuries B.C.E.) and to the investigation of harmonic intervals in terms of arithmetic. In classical Greece, harmonics was the science that investigated the arrangements of pitched sounds that form the basis of musical melody and the principles that govern them. What the early Pythagoreans discovered is that some differences in pitch between two sounds, or intervals, were expressible by numerical ratios (e.g., they found that the three basic intervals of octave, fifth, and fourth were expressible as the numerical ratios 2:1, 3:2, and 4:3). Thus, an arithmetical ratio could be successfully used to represent a perceptible phenomenon. To account for this fact, they proposed a metaphysical system wherein everything is a number understood via discrete representations. In this "all is number" doctrine, which is illustrated in Aristotle's *Metaphysics*, the world consists of corporeal numbers, or units, and mathematics applies to the world because the world has such mathematical structure.

Although the Pythagoreans pioneered the study of our topic, later attempts to explain the applicability of mathematics in science departed from their view. The Greek philosophers Plato (ca. 429–347 B.C.E.) and Aristotle (384–322 B.C.E.), for instance, developed different approaches to the applicability issue.

For Plato, the physical world and all of its constituents are imperfect copies, approximations, of special abstract entities called *Forms*. The world and its constituents derive from Forms, which include—but are not limited to—mathematical entities and properties. For instance, a drawn square belongs to the material world, but it is an imperfect, approximate copy of the square as such, whose angles are exactly 90° and whose sides are perfectly straight. Because all copies are dependent on ideal Forms, the physical world is dependent on Forms. Therefore, for Plato, mathematics is applicable to the physical world because it is a science of the ideal Forms from which the physical world is derived.

Aristotle disagreed with Plato about the nature of mathematical objects. While for Plato material objects are imperfect copies of ideal forms, and the former are derived from the latter, Aristotle maintains that mathematical objects are obtained from physical objects through a process of *abstraction*: We consider a physical object with respect to some of its attributes, next we take them out, and finally we create a new (abstract) object consisting in that object in only those respects. The result of this process is an object in its own right. The *snub-nose* is Aristotle's favorite example: If we consider a snub nose, the quality of snubness comes from sense perceptions and is concavity in the physical object *nose*; the mathematical quality of concavity, on the other hand, is obtained by removing all the attributes of the nose with the exception of its shape. Thus, according to Aristotle, mathematics studies the objects formed by abstraction but not the material objects on which the abstraction process operates. And this view provides an account of how mathematics applies to the world.

The philosophical standpoints proposed by Plato and Aristotle were very influential until the early modern period. The Italian astronomer, engineer, and physicist Galileo Galilei (1564–1642), for instance, echoes both views in *The Assayer* (1623) and in the *Dialogue Concerning the Two Chief World Systems* (1632). But the mathematization of physics that took place during the 17th century, largely influenced by the publication of Isaac Newton's *Principia* (1687), marked a radical departure from the dominant tradition of a natural philosophy that aimed to give qualitative explanations of natural phenomena. In the new conceptual framework that arose, where the objects of physics also included objects that were themselves mathematical (e.g., Newton's concept of force), philosophical analysis such as those offered by the Pythagoreans, Plato, and Aristotle were insufficient to account for the applicability of mathematics. Therefore, philosophers had to elaborate new strategies to explain the applicability of mathematics. Immanuel Kant (1724–1804), for instance, proposed—within his transcendental idealism—the idea that it is the constitution of the subject that makes it necessary to describe the world mathematically. In his *System of Logic* (1843), John Stuart Mill argued that mathematics is itself empirical since mathematical truths are generalizations derived from experience. Other philosophical standpoints were offered. Nevertheless, as happened with the accounts proposed by the ancient Greeks, these conceptions had to be revised or even abandoned in light of new, unexpected interactions between mathematics and science (e.g., the application of non-Euclidean geometry in Albert Einstein's theory of general relativity undermined Kant's view, and particularly his conception of the synthetic *a priori* character of geometry).

Successful interactions between mathematics and the empirical sciences have been growing massively since the first half of the 20th century. This phenomenon captured the attention of many scientists, as for instance the Nobel laureates in physics Eugene Wigner and Richard Feynman, and gave a new impetus to the philosophical study of the applicability of mathematics. Several analyses were developed but, in general, we can identify three main strategies that have been adopted to address the applicability issue: (1) a logicist approach, (2) an attempt to explain the applicability of mathematics by focusing on the way in which mathematical concepts are introduced in scientific theory-building, and (3) a strategy that elaborates on the idea that the applicability of mathematics can be explained in terms of some correspondences that can be established between a mathematical domain and an empirical domain. Let us consider the three strategies in turn:

1. In philosophy of mathematics, *logicism* is the view that mathematics, or part of it, can be reduced to logic. This view was developed by the philosopher Gottlob Frege at the end of the 19th century. According to the logicist standpoint, mathematics can be applied to the description of the world because it contains only logical concepts, which can be applied to any possible domain of objects. Although Frege's attempt to demonstrate logicism was undermined (because of an inconsistency that was found in his formal logical system), by the mid-1960s, some philosophers of mathematics tried to revive some of the core ideas of logicism and set in motion a philosophical movement that has been called *neologicism*. Neologicists defend *Frege's constraint* (also known as *application constraint*), a principle that was one of the main tenets of Frege's logicism and which requires that the empirical applications of a mathematical theory (or class of numbers) must be built directly into the formal characterization of the mathematical theory (or class of numbers). Thus, although in a renewed form, neologicist projects that meet Frege's constraint continue to offer an account of the applicability of mathematics by demanding that an explanation of applicability be provided inside mathematics, by mathematical definitions.
2. A second strategy, which was largely inspired by the emergence in the 1960s of a tradition in philosophy of mathematics that was concerned with the dynamics of mathematical discovery and the historical development of mathematics, focuses on the way in which the introduction of mathematical concepts is influenced by theory-building in science. Philosophers who follow this route draw on analyses of the historical development of mathematics and present case studies in which mathematics and science simultaneously develop in a theory-building context. By emphasizing the interplay between mathematics and other kinds of practices that we find in theory-building (e.g., the practice of identifying analogies between different scientific theories), they argue that the effectiveness of mathematics in science is no longer surprising when we see that (and how) much mathematics is brought into being by the need to model one or more aspects of the world (an example of such an analysis can be found in the paper "Solving Wigner's mystery," written by the historian of mathematics and logic Ivor Grattan-Guinness and published in 2008).
3. A third strategy, which has been the most influential since the first decade of the 21st century, gives an explanation of the applicability of mathematics in terms of mappings that are established between mathematics and the empirical systems we want to study. These mappings are mathematical and include not only isomorphisms but also other kinds of mathematical mappings like homomorphisms, epimorphisms, and monomorphisms. Mappings ensure that some crucial features of the empirical system are mirrored in the mathematical model used to study that system. Such an approach has been called *the mapping account of applied mathematics*. According to this view, the applicability of mathematics is fully accounted for by appreciating the relevant structural similarities between the empirical system under study and the mathematics used in the investigation of that system.

There have been many attempts to implement the mapping account, but the most discussed has been that proposed by Otávio Bueno and Mark Colyvan in their 2011 paper "An inferential conception of the application of mathematics." Bueno and Colyvan retain the basic idea behind the mapping account (i.e., the idea that the successful application of mathematics amounts to establishing a suitable mapping between mathematics and an empirical setup) and propose a view of applicability, called *the inferential conception*, in terms of a three-step process: (a) immersion (a correspondence between mathematics and the empirical setup is established via a suitable mapping), (b) derivation (some consequences are generated from the mathematical formalism), and (c) interpretation (the mathematical consequences obtained in the derivation step are interpreted in terms of the initial empirical setup, via a mapping that does not necessarily coincide with the mapping used in the immersion step).

Connections With Other Debates

The growth in successful interactions between mathematics and science played a motivating role in advancing analysis of the applicability issue. Nevertheless, philosophical interest in the applicability of mathematics was also prompted by the investigation of three topics that have become increasingly central to the agenda of philosophers of science and mathematics since the 1970s: mathematical modeling in science, mathematical explanation in science, and indispensability argument(s) for mathematical Platonism.

Mathematical Modeling in Science

Mathematical models are used in science to represent and study a selected part or aspect of the world, which is named the (model's) *target system*. For instance, the so-called Lotka–Volterra model, developed by the mathematicians Alfred Lotka and Vito Volterra in the 1920s and consisting in a pair of differential equations, is used in population ecology to represent a system of two interacting species (a predator and a prey) and study its dynamics. Sometimes, a mathematical model can also be used to represent and study two different target systems (e.g., the same equation can be used to represent and study a swinging pendulum and an oscillating electric circuit). Furthermore, the target system represented in the model is usually an *idealized* version of the real system, namely a simplified or distorted version of it. For instance, when we study a swinging pendulum, we usually assume that the pendulum bob is a point mass and that the string is massless. With these idealizations in place, the target system becomes more tractable from a mathematical point of view.

It is therefore easy to see how the investigation of the applicability of mathematics is closely connected to discussions of mathematical modeling in science: An account of how a mathematical model can be used to represent and study a target system prompts an explanation of how mathematics is applied with success to the physical world (or, if we consider the case of idealizations, an explanation of how mathematics is applied with success to those idealizations that are about the physical world).

Mathematical Explanation in Science

To explain something is to give the reason why that something happens (rather than a mere description of it). For instance, to explain a solar eclipse is to give an answer to the question "Why did the solar eclipse occur?" rather than a simple description of the eclipse phenomenon. Explanations are therefore particularly important in science, which investigates the reasons why empirical phenomena (such as the solar eclipse) occur.

Many explanations in science appeal to causes. These *causal explanations* provide information about what caused the thing that we want to explain (e.g., a solar eclipse is caused by the interposition of the moon between the sun and the earth). Moreover, many scientists and philosophers have argued that there also exist mathematical (and therefore noncausal) explanations in science. In this kind of explanation, it is mathematics that unveils the reason why an empirical phenomenon occurs. Nevertheless, there could be no mathematical explanation of empirical phenomena if mathematics were not successfully applied in the sciences dealing with such phenomena (note that the converse of this conditional statement is not true since there may be successful applications of mathematics that are not explanations). Therefore, articulating a plausible account of the applicability of mathematics in the empirical sciences is particularly important for those philosophers who maintain that mathematics can play an explanatory role in science.

Indispensability Argument(s) for Mathematical Platonism

Platonism in philosophy of mathematics is the view that mathematical objects exist (independently of us and our language, thought, and practices). The indispensability argument is a philosophical argument used by Platonists to defend their view and argue that there are mathematical objects. There exist several versions of the argument but, in its original formulation (which is credited to the philosophers Willard Van Orman Quine and Hilary Putnam), the argument goes as follows: Since we ought to believe in the existence of those entities that are indispensable to our best scientific theories (like electrons or black holes), and since mathematical entities are indispensable to our best scientific theories (i.e., our best scientific theories like quantum mechanics and general relativity cannot be formulated without making reference to mathematical entities), then we ought to believe in the existence of mathematical entities. In this line of argument, mathematical entities are seen to be on an epistemic par with other theoretical entities of science like electrons or black holes (and so anyone who believes in the existence of the former should also believe in the existence of the latter). Furthermore, the argument crucially presupposes that mathematics successfully applies in science. Hence, again, we see how investigations into the applicability of mathematics are linked to, and triggered by, the study of a separate topic.

Although discussing other formulations of the indispensability argument is beyond the scope of this entry, it is worth noting that there is a version of the argument that relies on the notion of mathematical explanation in science. Early presentations of such an *enhanced* (or *explanatory*) indispensability argument can be found in philosophy of mathematics before the end of the 20th century, but an explicit formulation appears in 2005 in Alan Baker's paper "Are there genuine mathematical explanations of physical phenomena?" Since, as discussed earlier in this entry, successful applicability of mathematics is a necessary—although not sufficient—condition for mathematical explanation in science, the issue of explaining why mathematics applies with success in science also arises in the context of the enhanced indispensability argument.

Daniele Molinini

See also Explanation; Inference to the Best Explanation; Laws (Scientific); Mathematics, 20th Century; Modeling; Philosophy of Science; Physics, 20th Century; Realism in Mathematics

Further Readings

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind*, *114*(454), 223–238. https://doi.org/10.1093/mind/fzi223

Bangu, S. (2012). *The applicability of mathematics in science: Indispensability and ontology*. Palgrave Macmillan.

Bueno, O., & Colyvan, M. (2011). An inferential conception of the application of mathematics. *Noûs, 45*(2), 345–374. https://doi.org/10.1111/j.1468-0068.2010.00772.x

Bueno, O., & French, S. (2018). *Applying mathematics: Immersion, inference, interpretation*. Oxford University Press.

Gingras, Y. (2001). What did mathematics do to physics? *History of Science*, *39*(4), 383–416. https://doi.org/10.1177%2F007327530103900401

Grattan-Guinness, I. (2008). Solving Wigner's mystery: The reasonable (though perhaps limited) effectiveness of mathematics in the natural sciences. *The Mathematical Intelligencer, 30*(3), 7–17. https://doi.org/10.1007/BF02985373

Mancosu, P. (2018). Explanation in mathematics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu

Panza, M., & Molinini, D. (Forthcoming). L'applicabilité des mathématiques. In P. Wagner, M. Panza, F. Poggiolesi, & A. Arana (Eds.), *Précis de philosophie des mathématiques et de la logique*. Editions de la Sorbonne.

Pincock, C. (2004). A new perspective on the problem of applying mathematics. *Philosophia Mathematica*, *12*(2), 135–161.

Steiner, M. (1998). *The applicability of mathematics as a philosophical problem*. Harvard University Press.

Urquhart, A. (2008). Mathematics and physics: Strategies of assimilation. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 588–620). Oxford University Press.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics*, *13*(1), 1–14.

Artificial Intelligence

Artificial intelligence (AI) is loosely defined as the capacity of machines that can accomplish tasks that humans would accomplish through thinking. This definition does not say anything about AI achieving this performance in ways similar to how we humans do it, however. The term *artificial intelligence* is used with two meanings. On the one hand, it refers to (artificially) intelligent machines and the ways of making them; in this sense, AI is primarily computer science and engineering. On the other hand, AI is also a transdisciplinary field of studying these machines. AI gurus such as Herbert Simon often emphasized that studying AI involves studying the human mind, and if we get it right, we will understand both AI and the human mind better in the end. Therefore, the field of AI involves various branches of hard sciences and engineering but beyond these also biology, psychology, and philosophy. This entry introduces the types of AI and describes their origins and the ways in which the types differ from each other. Next, the entry pursues the question of what defines intelligence and whether AI machines can be said to *think*. The entry continues with a discussion of the two paradigms of AI research, *strong AI* and *weak AI*. This is followed by an examination of the nature of knowledge and learning and the differences between AI and the human mind—a distinction reinforced in concluding remarks on the importance of this essential consideration in shaping the future of humankind.

Types of AI

The term *artificial intelligence* was coined by the computer scientist and cognitive scientist John McCarthy during a 2-month workshop organized at Dartmouth College in summer 1956. At the time, only one program (discussed below) qualified for the name, so it was the result of philosophizing about what computers should be capable of. To unpack the concept of AI, it is useful to distinguish between the different types of AI and to delineate facts from beliefs in the process. In this section, the three main types of AI, namely *symbolic reasoning systems*, *symbolic expert systems*, and *artificial neural networks (ANNs)*, are examined.

Symbolic Reasoning Systems

The first *thinking machine*—that is, the first AI implementation that worked—was called the *Logic Theorist*, or *Logic Theory Machine*. It was created in 1956 by the Nobel laureate Herbert Simon, his former PhD student Allen Newell, and Cliff Shaw from the RAND Corporation. The purpose of the Logic Theorist was "[. . .] to learn how it is possible to solve difficult problems such as proving mathematical theorems, discovering scientific laws from data, playing chess, or understanding the meaning of English prose" (Newell et al., 1963, p. 109). The Logic Theory Machine gradually evolved into the even more ambitious General Problem Solver, which was supposed to extend the original scope to all areas of human problem-solving. This was the first tentative attempt at what is today called *artificial general intelligence* (*AGI*).

The departure point was the philosopher Thomas Hobbes's idea from the 17th century, popularized by AI pioneer Oliver Selfridge (1926–2008), suggesting that all human problem-solving could be represented as a manipulation of symbols. If this was true, Newell and Simon speculated that computers should be able to solve real-world problems, not only arithmetic problems. They tried to achieve this by recording the steps of human problem-solving, using the *thinking aloud* technique: They asked problem solvers to note all the steps of their thinking. Newell and Simon figured that if they repeated the same in many areas of human problem-solving, they should be able to extract the general principles of reasoning, and by applying these principles and generic steps, AI should be able to work across multiple domains. The idea was a radical departure from the optimization and operational research methods of the time, which started with a detailed model of the complete problem. The performance delivered by the Logic Theorist was astonishing; eventually, it proved 38 of the first 52 theorems in the *Principia Mathematica*, a foundational work in modern symbolic logic by Alfred North Whitehead and Bertrand Russell.

What are the facts and what are the beliefs of symbolic reasoning systems? The first problematic assumption is that all thinking can be represented as symbol manipulation; there is no convincing evidence, so it is just a belief. The second assumption is that there are general principles that apply across all areas of problem-solving; it is difficult to imagine that there are some generic steps that apply from cooking a stew to designing a car or composing music. However, even if this does sound convincing, it is a mere belief. The third assumption is that playing chess or proving theorems, which may arguably involve explicit logical steps, has anything in common with "understanding the meaning of English prose" or anything that involves the achievement of meaning. It is nothing more than a belief. The fact is that the performance of the Logic Theorist was impressive, and since then, similar AI solutions have proved many more theorems and suggested new ones. What has not happened is an AI finding a surprising or significant theorem.

Symbolic Expert Systems

Often called the "father of symbolic expert systems" is Edward Feigenbaum (b. 1936), previously Simon's PhD student. He abandoned the general principles of reasoning to focus on narrow domains of expertise. He figured that in order to deliver performance on par with a high level of expertise, AI would need to have an internal *model*; a representation of the problem, which would be reasoned. As he began working with scientists, Feigenbaum anchored his project in experimental design in order to make testing possible. Joined by the Nobel laureate Joshua Lederberg, he decided to build a system that would induce molecular topologies from mass spectra, in order to support the Mars probe looking for life, or the precursors of life, on Mars. The project was named DENDRAL. The input data were reliable, the calculations needed were beyond the usual but not beyond the AI capabilities of the time, and there was a chance to test the findings empirically. The DENDRAL ran for many years, involving an increasing number of experts, eventually covering a larger area than any of the individual experts did. Eventually, the program was performing on par with, and exceeding, top experts.

The DENDRAL was the first *knowledge-based expert system*. A system being *knowledge-based* refers to the internal model, called *knowledge representation*, which is stored in the form of a *knowledge base* (this can be thought of as a cleverly designed database that stores knowledge), while the term *expert system* signifies, not surprisingly, that expert knowledge was represented. The knowledge representation is obtained in the process of *knowledge acquisition*, which is a subset of the overall process of *knowledge engineering*. The facilitator who works with the experts to model their knowledge is called the *knowledge engineer*. As DENDRAL was expanded, the complexity of the knowledge base increased, eventually threatening the stability of the initial LISP (a programming language focused on linked lists) system. Therefore, Feigenbaum's colleague Bruce G. Buchanan reprogrammed the system, establishing what became the standard for knowledge bases: the "production rules," or hierarchically organized "if . . . then" rules.

The facts about expert systems are that Feigenbaum was meticulous in selecting a suitable problem for this type of AI: The problem was of the right size, the right level of complexity, experts were available, and there was a possibility of checking the outcome. This is still the crucial feature of expert systems today. In other words, Feigenbaum's assumptions stood the test of time. Feigenbaum mentioned his belief in AI superintelligence, but this was not used as an assumption for developing expert systems. For many years, expert systems were *isolated*, in the sense that there was no connection and interoperability between knowledge bases; today, knowledge bases can be connected and provide input to each other. The main limit of expert systems today is *brittleness*, meaning that although expert systems deliver high performance in their narrow domains, they are completely useless even a tiny step beyond their domain's boundary, as there is no representation of nonspecialist knowledge.

Knowledge-based expert systems dominated the AI landscape until the mid-1980s, producing a large number of successful implementations in a wide range of areas from science and manufacturing through to medical treatments.

Artificial Neural Networks

The first ANN was conceptualized in 1943 by Warren McCulloch and Walter Pitts, while the first implementation, the SNARC (Stochastic Neural Analog Reinforcement Calculator), was created by Marvin Minsky and Dean Edmonds as part of their undergraduate work at Princeton University in 1950. The early ANNs were used to calculate mathematical functions; therefore, they should not be regarded as AI but rather as *pre-AI*. The popularity of ANNs rose in the mid-1980s, when computers became sufficiently fast to cope with ANNs of a useful size. Today, ANNs are the most often mentioned forms of AI; they occupy much of the AI landscape, including machine learning.

Any ANN consists of three sets of artificial neurons: There is an input layer, which receives a signal (stimulus); then, there is a hidden layer, which performs a translation or *black-box* operation; and finally, the output layer, which produces a response. The artificial neurons are connected to each other across the layers, resembling synapses in a brain. The response is compared with the expectations, the initial weights are amended, the process is repeated iteratively, and ANN will relatively quickly adjust and produce the desired response. In other words, ANNs learn from a large number of learning examples how to reproduce their statistical frequency. The often-mentioned *deep ANN* and *deep learning* simply mean that there is more than one layer of artificial neurons in the hidden layer.

When it comes to facts and beliefs, it is useful to consider how similar ANN is to the brain. The size of the human brain is estimated at around 80 to 100 billion neurons, each with around 7,000 connections on average, resulting in about 700 to 1,000 trillion (10^{15}) synapses in total. The largest ANN today consists of some 16 million artificial neurons, which roughly corresponds to the size of a frog's brain. Thus, in terms of synapses, we may be about 8 to 10 orders of magnitude short of the size of the human brain. At the same time, an ANN of 16 million neurons is already so large that the architect cannot grasp it anymore, and it takes a long time to train even using supercomputers. So, if it is even possible to approach the size of the human brain, this will not happen any time soon.

Structurally, ANNs are layered, and the connections usually only go forward (in some networks, there are also within-layer connections). The human brain is a complex network of neurons, so the idea of layered structure may not apply, and there may be circular connections. Functionally, an ANN reflects what was known about biological neurons around the mid-20th century: At any time, neurons are either *firing* or not. Today, it seems that at least a small subset of neurons (perhaps a few million?) display more complex behavior; that is, the strength of the firing impulse seems to matter. Individual neurons also display very different behaviors in this respect, and the strength of firing impulse can vary two orders of magnitude. All this means that structurally and functionally, ANN is rather limited in comparison with the human brain.

However, if ANN architects managed to construct something that resembles an artificial brain in terms of size, structure, and function, there is another belief that must be confronted: that the artificial brain would produce an artificial mind. What is known so far is that brain and mind are somehow connected, and particular thinking processes can be associated with activity in particular brain areas. We still lack an understanding of how the brain and the mind are linked.

Finally, according to recent research besides the brain, other organs, for example, the endocrine system, may be part of our cognitive system. It is not impossible that the mind is fully *embodied*, meaning that the whole body is part of it. So, an artificial brain may or may not produce an artificial mind.

Can AI Think?

Since the very beginning of AI, the big question has been: How do we know if AI should be considered (artificially) intelligent? The first criterion, proposed by Alan Turing in 1950, is still the most popular today. The essence of the Turing test is that if we communicate with an entity and cannot figure out whether it is a person or a machine, and in fact it is a machine, it means that the machine *thinks*, and it should be considered intelligent. At first sight, this argument may sound convincing; if the machine is not really thinking, we should recognize this. Not everything is, however, as it seems. The first program that passed the Turing test was Joseph Weizenbaum's ELIZA, created in 1966, followed by many other programs, up to today's chatbots. However, these programs passed the test with people who did not know that they were testing. The first experiment that was actually set up as a test was Eugene Goostman, the simulated 13-year-old Ukrainian boy. The legitimacy of this test can be questioned, but, as noted by Stuart Russell and Peter Norvig, it fooled 33% of the untrained amateur judges. So, the judges might not be right and 33% is quite low. In addition, the setup certainly worked in favor of the outcome: having someone who is 13, from a country about which the testers do not know much, speaking English as a foreign language. As Russell and Norvig speculated, perhaps the Turing test is really a test of human gullibility.

It is possible, however, that an AI will at some point legitimately pass the Turing test—but there are further problems. Turing modeled his test after the *imitation game*. In this game, Person ***A*** pretends to be Person ***B***. If ***A*** has learned a lot about ***B***, the testers may believe that ***A*** is ***B***. However, this does not mean that ***A*** actually became ***B***. The programs that passed the Turing test were designed with the purpose to pass the test, not with the purpose to think.

An excellent explanation of the Turing test is the *Chinese room argument*, proposed by the philosopher John Searle. The essence of the argument is the following: A person who does not speak Chinese is in a room full of rule books, receiving messages in the form of Chinese symbols and looking up responses to those messages in the rule books. If the rule books are good, those reading the responses outside the room may believe that the person responding speaks Chinese. Similarly, passing the Turing test does not prove that the computer thinks, only that the program is good.

Of course, none of these instances prove that machines cannot think; they only point out the lack of conclusive proof. So, we are left to our respective beliefs.

AI Paradigms

The distinction between *strong AI* and *weak AI*, as the two paradigms of AI research, was introduced by Searle, the author of the Chinese Room Argument. The strong AI paradigm postulates the idea of the thinking machine—of AI with a mind of its own, at least—as a desirable and achievable outcome. Most followers of strong AI do not believe that AI can currently actually think, but they do believe that it is only a matter of time until it will. Typically, this is fueled by the beliefs described in the previous and following sections of this entry. Marvin Minsky, a great proponent of strong AI, suggested that any endeavors in the field of AI, other than trying to create AGI, are meaningless.

In contrast, those who follow the weak AI paradigm see AI as a technology that can help produce useful and powerful tools. Many of them do not believe in the possibility of AGI, and none of them work on the AGI project. They argue that we cannot produce an artificial mind, as it would require the researchers to obtain a complete understanding of the mind, which they believe to be impossible. There is a very important consequence of this belief on what and how the followers of weak AI try to achieve in their work. They do not try to replicate the human thinking process; instead, they try to find a sensible way of achieving the same or similar performance. A popular metaphor used to describe this is the relationship between the horse and the car. When the original creators dreamed up the first automobile, they did not attempt to replicate the muscles, joints, and metabolism of a horse; they used an internal combustion engine instead, and put it on wheels. The car can be an excellent replacement for some aspects of the horse; for instance, it can take passengers from one place to another (more passengers than a horse, more conveniently and usually faster). In other areas, however, such as playing polo or enjoying horseback riding, the car cannot deliver. Similarly, what weak AI researchers produce only works in a narrow area and in very different ways from how humans would deal with the task.

Curiously, although their beliefs are fundamentally different and they focus on different types of work in the area of AI, the followers of the two paradigms do agree about one thing: that what is missing to achieve a thinking machine is common sense. As all human beings have common sense, it is sometimes considered to be something simple—but it is not. According to Minsky (1988, p. 22), common sense consists of "hard-earned practical ideas—of multitudes of life-learned rules and exceptions, dispositions and tendencies, balances and checks." It only appears simple because we do not recall acquiring it.

AI and the Human Mind

There are several areas that people think of as particularly human, such as knowledge, learning, and creativity. It is useful to look at the differences between the human mind and AI along these dimensions that can be conclusively demonstrated.

The difference that is the easiest to identify along the knowledge dimension is that systems that rely on knowledge acquisition are necessarily limited to explicit knowledge. According to the Hungarian–British polymath Michael Polányi, all human knowledge is either tacit or rooted in that which is tacit, as all explicit knowledge relies on being tacitly understood and applied. The problem of common sense showcased in the previous section can be explained similarly: Common sense is predominantly tacit, and all human knowledge incorporates common sense. *Intuition*, which has been deemed valuable in relation to peak human achievements (such as scientific discovery, artistic creativity, and managerial decisions), also belongs to the tacit realm. Humans also know *harmony and beauty*, and this beauty usually drives creative achievements, deep thoughts, and simply provides joy. For instance, a human can admire the beauty of a rainbow despite knowing that it is the result of the prism effect. People also use *analogical thinking*, which refers to mental models analogical to reality. An extraordinary example of this is the Serbian–American inventor Nikola Tesla (1856–1943) designing his machines entirely in his mind, without using a blueprint. Finally, *seeing the essence* is an achievement unique to the high levels of human expertise. It consists of two parts: the *big picture* and the detail. Those who achieved a high level of expertise are known to see the *big picture*. However, they can also see any of the details, as well as the relationship between the details as well as the relationship of the details to the big picture, and they can rapidly switch from one to the other.

Besides the many aspects of knowledge that do not seem to be easy to replicate or emulate with AI, knowledge also relates to other aspects of our being. Humans also have *feelings*, which are manifestations of biological drives such as hunger, fear, longing for the company of other people, and so on. While humans can temporarily override their feelings, using their knowledge—for example, we do not need to eat right away when we start feeling hungry—*emotions*, in turn, can override knowledge. Our important decisions, such as whom to marry, where to live, or what organization to work for, have significant emotional components. Minsky goes so far as to suggest that knowledge is impossible without emotions. Humans also have *values* embedded in their knowledge. For instance, sometimes they do not do what matches their well understood preferences but what they think is the right thing to do.

The situation with learning is similar to the situation with knowledge. Reinforcement learning, which is how ANN learns, is only a tiny part of human learning. There was a time when psychologists thought that all human learning is a variant of reinforcement learning, but psychology has come a long way over the past 70 years. The significance of talent has been discovered, the significance of which is that people learn faster in areas in which they are talented, and learning in these areas feels mostly like playing. AI learns equally in all areas. People get inspired and interested in various things, and inspiration and interest can make them invest extra efforts in learning what they like. Furthermore, those who achieved the highest level of expertise all went through some form of master–apprentice relationship. All people have some such experience; for instance, this is how we learn our native language. This immensely complex form of learning builds on a deep personal connection between master and apprentice, and our understanding of how it works is very limited. What we do know, however, is that the master–apprentice relationship seems to be the only way of transferring tacit knowledge.

Finally, there is a question if AI can be creative or not. It certainly outperformed top human experts in areas where creativity is considered important. Examples are Deep Blue winning against chess grandmaster Garry Kasparov in 1997, and AlphaGo defeating Lee Sedol in 2016 (as well as all the other Go grandmasters shortly after that). Human creativity is defined as the production of a new and useful idea—AI has definitely produced new and useful things, but it can be argued that it has never produced an idea. However, AI can certainly support human creativity, producing patterns from large amounts of data as well as helping us transcend the boundaries of our knowledge traditions.

Concluding Remarks: How to Think and Talk About AI

This entry has described what AI can do and what its limitations are; the entry has also delineated the facts and beliefs regarding current performance and future promises. AI comprises an exceptionally powerful set of technologies and may well be the most expensive human enterprise ever. If we use it well, it can greatly benefit the whole of humankind.

What can be done to get things right about AI? Perhaps it would be a good start to stop describing AI in human terms: *Human–AI collaboration*, *thinking*, *learning*, and *making decisions* are all terms that attribute intentionality to AI—something that it most definitely does not currently have, and possibly never will. Even when we talk about authentic AI, we usually think of a better fake human. Authentic AI, however, is not human-like; it is AI-like. For a human future in which AI plays a great role to the benefit of humankind, we need authentic humans as well as authentic AI.

Viktor Dörfler

See also Analysis; Cognitive Science; Intuition; Knowledge; Philosophy of Mind; Rationality; Reasoning

Further Readings

Davenport, T. H. (2018). *The AI advantage: How to put the artificial intelligence revolution to work*. MIT Press.

Dörfler, V. (2018). *Big data, artificial intelligence and the human mind*. Paper presented at the TEDx University of Strathclyde, Glasgow, UK. https://youtu.be/KAXfo-cZ80Y

Dörfler, V. (2020). Artificial Intelligence. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (3rd ed., Vol. 1, pp. 57–64). Academic Press.

Dörfler, V. (2022). *What every CEO should know about AI*. Cambridge University Press.

Dreyfus, H. L., & Dreyfus, S. E. (2000). *Mind over machine: The power of human intuition and expertise in the era of the computer*. The Free Press (Original work published 1986).

Feigenbaum, E. A. (1992). *A personal view of expert systems: Looking back and looking ahead*. Stanford University. https://purl.stanford.edu/gr891tb5766

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence* (2nd ed.). A K Peters.

McGilchrist, I. (2019). *The master and his emissary: The divided brain and the making of the western world* (2nd ed.). Yale University Press.

Minsky, M. L. (1988). *The society of mind*. Simon & Schuster.

Newell, A., Shaw, J. C., & Simon, H. A. (1963). Empirical explorations with the logic theory machine: A case study in heuristics. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 109–133). McGraw-Hill.

Penrose, R. (2016). *The emperor's new mind: Concerning computers, minds and the laws of physics*. Oxford University Press (Original work published 1989).

Polányi, M. (2002). *Personal knowledge: Towards a post-critical philosophy*. Routledge (Original work published 1962).

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education.

Searle, J. R. (2004). *Mind: A brief introduction*. Oxford University Press.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). MIT Press.

Winograd, T., & Flores, F. (1987). *Understanding computers and cognition: A new foundation for design* (2nd ed.). Addison-Wesley.

Assumption

Our contemporary word *assumption* is derived from the Latin word referring to the act of being taken up or received. In that context, the Catholic Church used the word to articulate the Assumption of the Virgin Mary—that is, the official dogma that she was taken up bodily into Heaven. That use remains today, but beginning in the 13th century the word slowly took on the broader meaning of any idea, premise, or axiom that is *taken up* or taken for granted. With this latter meaning in mind, in what follows, a definition of assumption is offered and explained point by point, giving examples along the way as the term relates to both theory and human life. From there, a few additional examples are provided of the importance and place of assumptions in theory and theory building, spanning not only the physical sciences and mathematics but also the social sciences. The entry concludes with a brief summation of the place and importance of assumptions in theories.

What Is an Assumption?

Assumptions are beliefs of a particular kind—namely, assumptions are propositions that persons (and in the aggregate, human groups or even entire societies) regard to be true, whether consciously or tacitly, and which form the basis of—rather than being conclusively inferred from—processes of reasoning and lived experience. This definition requires some unpacking. First, assumptions are propositional, but oftentimes this is the case merely in the sense that, in principle, they could logically be articulated as propositional statements (unlike, for instance, some kinds of emotional states). In other words, assumptions are propositional not in the sense that they are necessarily or even usually propounded and spelled out in explicit propositional form. But assumptions are propositional in the sense that they make claims regarding what is true about reality and those understandings—even when left unstated—can be either true or false.

Secondly, assumptions are not free-floating ideas but are believed by persons and often also by groups, subcultures, or societies. It is not hard to grasp that persons hold fast to assumptions. But when persons gather, interact, or even share some common social identity from a distance, their association is likely to give rise to (or be premised upon) taken-for-granted ideas, values, and axioms. The first-generation college student group at the state university likely holds certain assumptions—about class mobility, equality, and the purpose of higher education—which take intentional work to see and to understand. It is also possible for the vast majority of a society to hold to an assumption that is rejected by a subcultural subset within it. For example, the popular assumption in Western modernity that what is morally right and good is ultimately up to each individual to decide is rejected by many conservative religious thinkers and some philosophers.

Third, assumptions can function anywhere along a spectrum from conscious to routinized to subconscious, being explicit or left implicit. In some cases, assumptions are stated up front and explicitly recognized as providing the grounds for some subsequent argument, observation, or experience. This is the case, for instance, in many philosophical arguments (especially those in analytic philosophy using formal logic) as well as in the mathematical proofs, one undertakes in algebra or geometry classrooms. In this sense, assumptions are rightly understood as taken-for-granted premises, the logical givens that set the stage for what comes next in the course of making a case or building an argument. In other cases, assumptions are known but nevertheless left unexpressed. For example, an evangelical Christian might try to organize her life and beliefs around the idea that the Bible alone is God's authoritative Word to humans. Such a baseline assumption would presumably have a powerful guiding effect on her life and beliefs even if most days it does not come up in conversation, but if asked she would be readily able to express it.

At still other times, assumptions operate on a level that is wholly unrecognized and implicit. Persons can hold assumptions without even knowing that they hold them, and persons can be powerfully influenced by those assumptions without recognizing it. Assumptions—as a particular kind of belief—need not be reflected upon, consciously adopted, or able to be stated in order to be genuinely believed. This does not mean, though, that such assumptions must always stay at the prediscursive, implicit level of consciousness. It is often possible for another person to point out and articulate a previously unrecognized assumption, at which time that assumption might move from unrecognized to recognized, from implicit to explicit. It is even possible for a person to identify her own previously unrecognized assumptions and change them if necessary.

Sometimes this is difficult, requiring study, dialogue, and critical reflection. Other times, this does not take much. Persons often do not realize they had been holding some assumption until it turns out to be false: You return home from a weeklong vacation only to discover that you had wrongly assumed that your spouse had locked the front door. You arrive at your favorite coffee shop to meet a friend, not having considered that she meant the one on the other side of town.

Fourth, assumptions are not (and cannot be) conclusively inferred or verified from reason or experience. Against the tradition of rationalism, taking assumptions seriously in life and in academic theorizing requires the recognition that reason, rationality, and logic are never the only factors at play. Likewise, against the tradition of empiricism, seriously accounting for the role of assumptions in life and theorizing necessitates moving beyond a strict focus only on what can be observed from experience and the five senses. Instead, assumptions are precommitments and premises that come before either reason or empirical experience. In the language of epistemology (the philosophy of knowledge), an assumption can be *justified* (i.e., a person is not always being cognitively reckless or intellectually sloppy for believing it) without being *justifiable* (i.e., without being able to be verified through argumentation and evidence). This does not mean that assumptions cannot be supported or made compelling through reasoning and experience. If an assumption meshes well with one's lived experience and helps to make the world intelligible, then there is good reason to hold to it. It is justified. But it cannot be conclusively established or deduced with indubitable certainty through the use of reason or observation. It is not justifiable.

Fifthly, basic assumptions form the grounds for processes of reasoning—ranging from particular instances of philosophical argumentation and mathematical proofs to more general approaches to (or convictions about) what knowledge is and how humans can go about acquiring it. This is most easily seen in simple logical syllogisms, in which a major premise and a minor premise are used to deduce a conclusion. But even in more complex belief structures and systems of reasoning, assumptions can serve the role of basic (and powerfully life-guiding) beliefs. That is, certain kinds of beliefs—such as that a divine being exists, that one's memories really do refer to actual past happenings, or that natural science is the only rational way to reach conclusions—can *sit at the bottom*, so to speak, of one's system of beliefs or worldview. Such basic beliefs are not inferred from, or held on the grounds of, other beliefs that the knower regard to be true. And different persons (and human groups) can and do hold to differing beliefs in this basic way. This pluralistic nature of basic starting points has led many philosophers of knowledge away from foundationalism (i.e., the theory that all humans share some common, incorrigible foundation for knowledge) to what is known as *post-foundationalism*. This is also why it might seem like a life-altering paradigm shift when one's basic beliefs and assumptions, for some reason, are revised.

Sixth, assumptions and premises are crucial not only for theory and theorizing but also for human experience and life in the world more broadly. The propositional content of assumptions can range from the mundane ("Oh, I assumed you were going to buy the milk") to big questions in the realms of epistemology, metaphysics, and ethics. And to the extent that abstract ethical, epistemological, and metaphysical issues and ways of thinking influence the ways persons live, even seemingly distant assumptions turn out to have serious consequences on actual experience. On this point, some philosophers and social scientists emphasize the principle of *practical adequacy*, which suggests that when the components of human knowledge systems (including one's assumptions) *rub up against* practical activity in the world in a way that makes the world and life unintelligible or unlivable, there is good reason to suspect that some crucial part of that knowledge system is problematic (i.e., false) and thus may need to be reconsidered and revised.

Examples of Assumptions

Assumptions play a central role in theories and theory building in several ways. In the social sciences, for example, assumptions about the nature (or lack thereof) of human beings are multiple and lead theorists in different directions. Some theorists hold a social collectivist view that claims human beings are ultimately a product of the cultural environments in which they are embedded, such that they are primarily concerned with living out their society's norms and values. Other theorists assume a fundamentally economistic view of humans as self-interested rational actors, seeing humans as driven by cost–benefit calculations in order to maximize their utility. Other theorists, influenced by postmodernism and phenomenology, view human identity and selfhood as fragmented, transient, and variably constructed across different situations—lacking any essential nature. Still other theorists adopt a sociobiological and evolutionary view of humanity, in which all of human social life can be explained by reference to *survival of the fittest* and the functioning of genetic material. And some theorists hold to an Aristotelian view of human beings as persons with an essential shared human nature and a moral direction toward which to strive for their objective flourishing. Which of these models of human nature is taken for granted clearly would have a great impact on a social scientist's explanations and theories.

Another example can be drawn from the history of geometry. In the third century B.C.E., the Greek mathematician Euclid systematized geometry based on five axioms, from which more complex theorems could be deduced. These five axioms involved the relationships between lines, points, and angles (e.g., between any two points there is only one line segment with those points as end points). For more than 2,000 years, this axiomatic system was the only version of geometry. But over the centuries, several thinkers questioned whether Euclid's fifth axiom (namely, that for any line and a given point off that line, there is only one line that runs through that point and parallel to the first line) was self-evidently true. Several attempts to prove the fifth axiom based on the first four ended in failure. Beginning in the early 1800s, mathematicians started to imagine what alternative geometric systems would look like if Euclid's fifth axiom—known as the *parallel postulate*—could be false. The result was a variety of non-Euclidean geometries, in which the spatial grid on which geometry is done is not a boxy, straight *Euclidean space*, but instead is intrinsically curved in various ways. This questioning of Euclid's fifth assumption and the recognition that spatial grids can be curved helped to shift the way physicists think about space and set the stage for Albert Einstein's theory of general relativity and gravitational force.

Even the scientific method operates on the basis of a handful of assumptions that cannot be verified using the scientific method. For example, science proceeds on the presupposition that there exists a real world about which scientists are trying to gain knowledge and that this world has a structure and nature independent of our theories about it. Scientific inquiry and discovery also makes sense only if scientists believe that aspects of the physical world (such as the laws of physics or the basic features of biology) will work very similarly tomorrow as they have worked in the recent past. Likewise, the scientific method assumes that human cognitive and sensory faculties are fairly reliable tools for gathering and interpreting data. Even more basically, science proceeds under the supposition that some beliefs and theories are closer to the truth than others; all is not relative. Science even makes moral assumptions—such as that theories should be tested fairly, that findings should be reported honestly, and that it is better to believe what is true than what is false. As assumptions, such beliefs are not necessarily unjustified or wrong, but they are nevertheless not definitively verifiable by logical proofs, arguments, or empirical observations.

Conclusion

It is not necessary—and probably not possible—for every unverified assumption to be articulated in theory and research. But clearly, assumptions are both important and unavoidable. While some scholars may feel the urge to hide, ignore, or deny the premises and assumptions that underpin their work, it is best as much as possible to be aware of the assumptions that one is making. And beyond just reflecting upon and recognizing one's assumptions, it is important as well to be willing to question and possibly to change them when the data and evidence call for it. Even the most fair-minded scholarship does not operate with a God's-eye view or with every idea or belief established with absolute certainty. But with sustained and critical attention to the assumptions undergirding scholarship, development of theory can proceed on a steady footing.

Brad Vermurlen

See also Belief Revision; Epistemology; Warrant; Worldview

Further Readings

Alston, W. P. (2005). *Beyond "Justification": Dimensions of epistemic evaluation*. Cornell University Press.
Llewelyn, J. E. (1962). Presuppositions, assumptions, and presumptions. *Theoria*, *28*(2), 158–172.
Schlefer, J. (2012). *The assumptions economists make*. Belknap Press.
Smith, C. (2003). *Moral, believing animals: Human personhood and culture*. Oxford University Press.
Ullman-Margalit, E. (1983). On presumption. *Journal of Philosophy*, *80*(3), 143–163.

Astronomy

See Cosmology

Authority

The concept of authority is fundamentally related to the concepts of hierarchy, power, and obedience. In modern research on authority, the concept has been articulated and operationalized in a number of different ways. Scientific discourse is discussed in terms of its influence and paradigmatic power, intrinsically linked to its authoritativeness. Political scientists refer to systems of authority and how they relate the role of the state. Psychologists refer to the relationship between authority and obedient behavior.

The word *authority* is etymologically rooted in Latin and is connected with the words *auctoritas* and *auctor*, which reference "one who originates or promotes ideas." In Old French, the term *autorité* is itself related to the Old English term for *author*, again referencing "one who originates." These roots are reflected in the general concept of authority as "one who originates or issues command or orders," ostensibly to be obeyed by those subservient to the person, their social role, and/or the orders. Conceptually speaking, these dynamics naturally invokes reference to the constructs of hierarchy and the power differentials that underlie and support a hierarchical social system. Philippe Aghion and Jean Tirole note the distinction between *formal* and *real* authority in organizations, the former being ascribed to an office who holds power by virtue of the position (e.g., a chief executive officer, or CEO, who may not actually make real decisions impacting organizational goals) and the latter having the ability to make decisions on the basis of authority granted by the superior.

This entry considers the varying conceptualization of authority, and how they have been reified in scientific, political, religious, and psychological narratives. This list is by no means exhaustive but is meant to synopsize some of the most prominent discourses that have been developed around explicating authority and how it structures human behavior.

Scientific Authority

Scientific discourse is powerful and influential to the extent that it is considered authoritative. Thomas Kuhn described the manner in which scientific theories come to dominate others and eventually become widely accepted as lenses through which problems are both conceptualized and solved. In his influential book *The Structure of Scientific Revolutions*, published in 1970, Kuhn describes the social process whereby scientific theories become authoritative, a point to which he ascribes the term *paradigm*. Scientific paradigms do change over time, largely due to the presentation of anomalies that cannot adequately be addressed or resolved within the context of the dominant paradigm of the time. When this happens—and Kuhn suggests that it is ultimately inevitable—a paradigm shift occurs whereby novel, unanticipated solutions to the anomalies arise, and achieve acceptance in the wider scientific community. Over time, the new paradigm comes to be accepted as authoritative among the scientific community and society in general.

Political Authority

Political authority is a concept that is intrinsically linked with the concepts of power and legitimacy. Power, in its most abstract conceptualization, typically entails the ability to influence others behavior beyond what they would ordinarily do, and legitimacy is most often conceptualized as a form of government which citizens believe is appropriate for that society. C. W. Cassinelli noted that legitimacy and political authority almost always involve an element of coercion, but can never involve force or the explicit threat of it; to do so is to undermine the extent to which one is truly considered a political authority.

Max Weber (1864–1920) developed a typology of political authority in his treatise "The Three Types of Legitimate Rule": traditional, legal-rational, and charismatic. Weber's three types of authority represent ideals, and thus do not necessarily correspond precisely with any given sociopolitical entity. In addition, authority can evolve from the lowest form into one of the other, more advanced, forms. *Charismatic authority* is described a social arrangement wherein leaders derive their power on the basis of their possession of some type of extraordinary attribute or supernatural power. Those subject to this type of rule must inherently believe in the power granted the leader. A contemporary example of a state predicated upon charismatic authority is North Korea under the totalitarian rule of the Kim family. The Kim family began their reign with Kim Il Sung, a leader whose persona was so elaborated by means of state-controlled propaganda, he continued to serve for years after his death as the world's only posthumous president.

Traditional authority arises from longstanding custom and is the basis of authority for monarchies. The death of a charismatic leader can result in the formation of a system whose authority derives from tradition. This type of authority is oriented toward maintenance of the status quo ante and its inherent inequalities. Beyond the domain of nation-states and their form of political organization, an example of this type of authority can be found in patriarchal societies, wherein the inherent inequality between men and women has been established not on the basis of charisma (or rational principles) but solely on the basis of longstanding custom.

The *legal-rational system of authority* is organized such that citizens are subservient to a legal structure that has been rationally constituted, rather than to a particular person, as is the case within a system predicated upon a charismatic or traditional form of authority. That legal structure can take the form of a constitution, a static form of bureaucratic organization, or rationally derived codified principles that establish the legality, responsibilities, and conditions of rule. Typically, under this form of authority, citizens are deferential to the office wherein power is vested, and not to the particular officeholder who happens to fulfill that social role. As an example, 21st-century Japanese, British, or Canadian political authority would be characterized as legal-rational in organization.

Religious Authority

Mark Chaves defines religious authority as "a social structure that seeks to enforce its order through the legitimate control of some supernatural component" (1994, p. 750); where authority withholds access to something vis-à-vis the legitimacy of the supernatural, that authority is religious. The goods that religious authority might control vary between religious traditions (e.g. eternal life, an end to suffering, or prosperity). Authoritarianism has been linked with religiosity and is a social attitude that refers to individuals who hold conservative principles and are generally accepting of established authorities. Authoritarians and deeply religious people tend to share common values, preferring the status quo over change, and employing a noninterpersonal morality that refers to external sources of authority (e.g., group norms).

Psychological Authority

The psychological literature has considered the nature of authority as it manifests itself in interpersonal relationships (such as between parent and child) and how authority is related to obedient behavior. The scope of discourse surrounding parental authority generally concerns parents' ability to utilize one of two types of control: psychological control and behavioral control. Psychological control involves any attempt by parents to encroach upon a child's psychological and emotional development. It typically takes the form of manipulative tactics in order to achieve its goal. Behavioral control refers to parental attempts at managing a child's behavior, such as correcting a child's inappropriate behavior with punishment.

Psychological control is most potent in young people before and during adolescence and is sometimes thought to be a parenting style as opposed to a parenting practice. This type of control does not necessarily involve parental pressure to feel and think in ways dictated by parents; it can merely embody parental pressure aimed at making the child behave in accordance with parental expectations. Tactics like guilt induction help parents maintain legitimate parental authority over issues of concern, like child friendship affiliations. Parental control can sometimes intrude upon a child's psychological development. Where it does, empirical evidence suggests that the negative consequences of this control can be predictive of later mental health issues and behavioral disorders. Where it concerns behavioral control, research on parental authority suggests that it is not counterproductive unless it impairs the bond between parent and child; harsh or erratic levels of punishment are predictive of later involvement in activities such as delinquent behavior.

Parents are encouraged to balance appropriate control over their children's behavior with developmentally appropriate attempts to give them autonomy as they grow older. Younger children are more accepting of parental authority and its legitimacy; however, some have found that as children aged, their views on parental authority become less positive. Where it concerns the scope of parental authority, there appear to be clear boundaries with regard to what constitutes legitimate parental authority and what should remain under the jurisdiction of children.

Stanley Milgram and *Obedience to Authority*

One of the most controversial and thought-provoking research projects aimed at understanding the nature of authority was psychologist Stanley Milgram's work at Yale University concerning obedience to authority figures. His research, first described in the *Journal of Abnormal and Social Psychology*, was published contemporaneously with Hannah Arendt's coverage in 1961 of the Adolf Eichmann trial for *The New Yorker* magazine, the hugely consequential trial of a former German bureaucrat instrumental in supervising, organizing, and facilitating the systematic extermination of European Jews, or *Jewish Affairs*, as it was euphemistically put in German bureaucratese. Fifteen years after the close of the war, Eichmann had been discovered living in Argentina by the Simon Wiesenthal organization; he was subsequently abducted by Mossad, the Israeli intelligence agency and smuggled back to Israel on a commercial flight to stand trial as personally accountable for his role in the murder of 6 million Jews.

Milgram was explicit about his interest in exploring the social dynamics that had facilitated the implementation of the Final Solution to the Jewish Question, the event now formally referred to as *the Holocaust*, and intended to compare levels of compliance in Germany with those in the United States. Aside from inspiring an onslaught of critique from research ethicists, his 1974 book *Obedience to Authority: An Experimental View* exposed a social dynamic whereby free will can be truncated, overridden, in the name of obedience to those who are perceived to constitute legitimate authorities.

The experimental design saw to it that experimental subjects administered an incrementally more intense series of electric punishment shocks to an ostensible *subject* (in reality a confederate of the experimenter) under the guise of its being a test of negative punishment's effect on learning capability. The coercive prompting of a confederate *researcher*, playing the role of a legitimate authority figure, insisted at appropriate times that subjects continue with the experiment despite any protests they might have to delivering increasingly strong electrical shocks to the victim, the bogus *subject*. Milgram performed many permutations of the relationship between the various actors involved in the experiment—altering in varying degrees the proximity of the victim, confederate researcher, and the experimental subject in order to test whether social distancing had an impact on a subject's obedience to orders, which was generally found to be the case.

Replications of the shock experiments were canceled in Hamburg, Germany, when Milgram found levels of compliance in the United States (the control group) high enough to rule out meaningful comparison with what he originally conjectured would be a more obedient German population. To the lament of the predictions of the psychiatrists Milgram had surveyed beforehand, roughly two-thirds (66%) of the subjects in Milgram's experiments were fully compliant in harming (and ostensibly killing) the confederate victim. The dynamic that Milgram exposed was one whereby personally injurious behavior need not be related to any moral, psychological, or biological factors impacting the experimental subject. Milgram's overall findings are compatible with the sincerity of the defense cliché espoused by those tried for offenses against humanity at Nuremberg and elsewhere: that they were merely *following the orders* of their superiors, presumably because they did not have the grounds to question the legitimacy of that authority, or the commands issued forth.

Douglas J. Dallier and Olivia M. Lewis

See also Evidence; Fact Versus Theory; Warrant

Further Readings

Aghion, P., & Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy, 105*(1), 1–29.

Asquith, P., & Smetana, J. G. (1994). Adolescents' and parents' conceptions of parental authority and personal autonomy. *Child Development*, *65*(4), 1147–1162.

Barber, B. K. (1996). Parental psychological control: Revisiting a neglected construct. *Child Development*, *67*(6), 3296–3319.

Cassinelli, C. W. (1961). Political authority: Its exercise and possession. *Political Research Quarterly*, *14*(3), 635–646.

Chaves, M. (1994). Secularization as declining religious authority. *Social Forces*, *72*(3), 749–774.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. University of Chicago Press.

Milgram. (1969). *Obedience to authority; An experimental view*. Harper & Row.

Weber, M. (1958). The three types of legitimate rule. *Berkeley Publications in Society and Institutions*, *4*(1), 1–11.

Axiom Schema

An *axiom schema* (plural: *schemata* or *schemas*, from the Greek *skhēma*, meaning *form* or *figure*) is a *template* for an axiom, accompanied by a *rule* that tells us how to instantiate the template with specific expressions in some (formal) language to produce an axiom. Thus, in developing an axiomatic theory in logic, mathematics, or science, one uses axiom schemas to specify a collection of axioms representing the basic or fundamental assumptions of the theory, from which one derives all other implications of the theory (i.e., its *theorems*). However, there will typically be *infinitely* many such axioms and so some method is needed to specify all of them at once by finite means. It is primarily because of this that we require the templates for axioms known as *axiom schemas*.

It is worth noting that *axiom* schemas are just one instance of the slightly broader notion of a schema, used in both logic and mathematics. The broader notion includes templates both for the axioms of a theory and so-called *rules of inference*, that is, rules for deducing consequences from the theory's axioms or from previously established theorems or previously made assumptions. Because there will also be an infinite number of ways of applying a given rule of inference, a schema for representing this rule by finite means is also needed. This entry discusses examples of axiom schemas in both logic and mathematics.

Axiom Schemas in Logic

Axiomatic systems in logic, often called *formal systems* or *logical calculi*, allow us to determine which statements in a formal language are logical validities, either because they are axioms themselves or because they are logical consequences of axioms. For example, in a *Hilbert system* for propositional logic, an approach that emphasizes axioms and minimizes the role of deductive inference rules, the following expressions are examples of axiom schemas:

$$\text{P1: } (\phi \rightarrow \phi)$$

$$\text{P2: } \big(\phi \rightarrow (\psi \rightarrow \phi)\big)$$

$$\text{P3: } \big((\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \chi))\big)$$

$$\text{P4: } \big((\sim(\phi) \rightarrow \sim(\psi)) \rightarrow (\psi \rightarrow \phi)\big)$$

In each of these four axiom schemas, we have expressions representing logical connectives; here the arrow, “ρ ,” represents material implication and the tilde “~” represents negation. Other systems might also involve symbols for disjunction (“∨”), conjunction (“∧”), and the biconditional (“↔”). Technically, the left and right parentheses are also part of this logical vocabulary, necessary for indicating and disambiguating the scope of the logical constants. We also have the expressions “ϕ,” “ψ,” and “χ” representing arbitrary well-formed formulae in the formal language in which the system is expressed. These latter expressions are sometimes called *metalinguistic variables* as they are not themselves expressions in the formal language of the system but are expressions from the *metalanguage*, the language we use to talk about and study the formal language and properties of the system. Thus, in contrast to the metalanguage, the language of the system itself is called the *object language*, and, in the context of propositional logic, the metalinguistic variables range over the class of well-formed formulae (or *wffs*) in the object language. Most commonly either Greek letters or uppercase Latin letters (*A*, *B*, *C*, . . .) are used to denote metalinguistic variables.

Each of P1 through P4 is thus a template in the sense of being a string of symbols that outlines and represents the shape or form that an axiom in the Hilbert system of propositional logic can take. However, these axiom schemas also implicitly involve a *rule* indicating how they are to be instantiated or *filled in* for the sake of producing a genuine axiom expressed in the language of the system itself. That rule is implicit in the range of values that the metalinguistic variables, “ϕ,” “ψ,” and “χ” can take, which is here stipulated to be the class of well-formed formulas of the formal language of propositional logic. To show what these instances might look like we can first briefly sketch a formal language for the propositional logic embodied in this Hilbert System:

Let L_{PROP} be the language of propositional logic, defined as the following set of wffs:

i. $P_i \in L_{PROP}$, for every $i \in \mathbb{N}$

ii. If $\phi \in L_{PROP}$ then $\sim(\phi) \in L_{PROP}$

iii. If $\phi, \psi \in L_{PROP}$ then $(\phi \to \chi) \in L_{PROP}$

iv. Only expressions formed by applications of (i)–(iii) are wffs of L_{PROP}.

Thus, applying these formation rules, the following strings of symbols are all examples of well-formed formulae in L_{PROP}:

$$a.\ P_1 \to (P_4 \to P_{23})$$

$$b.\sim\ (P_2 \to P_{217}) \to (P_{1300} \to (P_{14723} \to\sim\ (P_0)))$$

$$c.\ P_5 \to\ ((P_7 \to\ (\sim\ P_9 \to P_{11})))$$

Procedurally, having first specified a formal language for propositional logic so that we now know what counts as a wff, the axiom schemas P1 through P4 are then introduced to specify which specific wffs in the language count as axioms of the deductive system. The following are therefore axioms of the system as they are instances of the axiom schemas P1 through P4, respectively:

$$(P_7 \to P_7)$$

$$(P_7 \to (P_{10} \to P_7))$$

$$(P_2 \to (P_{111} \to P_{6455})) \to$$
$$((P_2 \to P_{111}) \to (P_2 \to P_{6455}))$$

$$((\sim P_5 \to \sim P_{108044}) \to (P_{108044} \to P_5))$$

The following more complicated formulae are also axioms and instances of P1 through P4, respectively:

$$(P_7 \to (P_{42} \to (P_{500} \to P_{99}))) \to$$
$$P_7 \to (P_{42} \to (P_{500} \to P_{99})))$$

$$(\sim (P_{80} \to P_7) \to ((P_{10} \to P_{778}) \to$$
$$\sim (P_{80} \to P_7)))$$

$$(P_2 \to (\sim (P_{111}) \to \sim (P_6 \to P_{89}))) \to$$
$$((P_2 \to \sim (P_{111})) \to (P_2 \to\sim (P_6 \to P_{89})))$$

$$((\sim (P_7 \to (P_{10} \to P_7)) \to \sim (P_{10} \to P_{778})) \to$$
$$((P_{10} \to P_{778}) \to (P_7 \to (P_{10} \to P_7))))$$

The axioms can thus become almost arbitrarily complex. In a Hilbert system for propositional logic, the axiom schemas are typically also supplemented with a schema for a rule of inference indicating how one formula may be permissibly transformed into another within the system. A common example of a rule used in this context is *modus ponens*:

$$\frac{\begin{array}{l}(\phi \rightarrow \psi)\\ \phi\end{array}}{\psi}$$

This rule tells us that from any pair of formulae that have the schematic forms represented by "$(\phi \rightarrow \psi)$" and "ϕ" we can infer a formula with the form represented by "ψ." Again, there will be infinitely many pairs of formulae that have this form, and so the modus ponens schema allows us to generalize over all of them by finite means.

In more powerful logical languages and systems, such as the first-order predicate calculus and axiomatic systems of first-order logic, axiom schemas are again used to specify an infinite collection of axioms by finite means. However, unlike with those of propositional logic, the templates and rules for axiom schemas in first-order logic also deal with expressions of types other than just the wffs in the language, in line with the greater expressive power of the predicate calculus. The following axiom schemas for the universal quantifier '∀' in first-order logic illustrate this difference:

$$\text{PC1: } (\forall\alpha(\phi) \rightarrow \phi(\beta\ /\ \alpha))$$

$$\text{PC2: } (\forall\alpha(\phi \rightarrow \psi) \rightarrow (\forall\alpha(\phi) \rightarrow \forall\alpha(\psi)))$$

$$\text{PC3: } (\phi \rightarrow \forall\alpha(\phi)) \text{ when "}\alpha\text{" is not free in } \phi$$

In PC1, the expression "$\phi(\beta/\alpha)$" indicates a metalinguistic variable for a wff that is exactly the same as that instantiating the metalinguistic variable "ϕ" *except* that it may differ (at most) in having the term that instantiates the metalinguistic variable "β" substituted for all occurrences of the bound term instantiating the metalinguistic variable "α." PC2 is more straightforward, indicating that the universal quantifier binding a term distributes over the material conditional. Finally, PC3 tells us that all instances of the conditional "$(\phi \rightarrow \forall\alpha(\phi))$" are axioms whenever the bound term instantiating the metalinguistic variable "α" does not occur unbound (*free*) in the wff instantiating the metalinguistic variable "ϕ." As such, we can see that axiom schemas for axiomatic systems of first-order logic require variables for expressions other than wffs, that is, they require variables for *terms* (where the terms include variables, constants, and function applications in the language of the predicate calculus). Relative to a first-order language with variable terms "x_0, x_1, x_3 ...," constant terms "c_0, c_2, c_3 ...," (unary) function terms "f_0, f_1, f_3 ...," and (monadic) predicate terms "F_0, F_1, F_3 ... ." Some possible instances of PC1–PC3, respectively, are the following:

$$(\forall x_0(F_1x_0 \rightarrow F_1x_0) \rightarrow (F_1c_5 \rightarrow F_1c_5))$$

$$(\forall x_7(F_2x_7 \rightarrow F_8x_7) \rightarrow (\forall x_7(F_2x_7) \rightarrow \forall x_7(F_8x_7)))$$

$$(F_4f_{88}(c_{17}) \rightarrow \forall x_3(F_4f_{88}(c_{17})))$$

As with the language of propositional logic, there will be infinitely many such axioms, most of which will be much more complicated than these three. The value of these axiom schemas is that we don't have to worry about the impossible task of specifying these axioms individually, since the schemas allow us to capture them all in one fell swoop.

Axiom Schemas in Mathematics

Moving from logic to mathematics, axiom schemas are also indispensable in the formal study of axiomatic theories. However, it is perhaps worth noting that the techniques and practices of most working mathematicians are typically less emphatically formal than those of logicians. Two well-known examples are worth considering, however. The first is from first-order Zermelo–Fraenkel set theory and is known as the *axiom schema of separation*. One possible formulation of this axiom schema is as follows:

$$\forall s_0(\exists s_1(\forall s_2((s_2 \in s_1) \leftrightarrow ((s_2 \in s_0) \wedge \phi))))$$

Here, the metalinguistic variable "ϕ" can be replaced by any formula in the language of first-order set theory. In practice, "ϕ" will typically be instantiated with a formula in which the variable "s_2" occurs free. Hence, the formula instantiating "ϕ" will represent some condition that the values of "s_2" must satisfy, thus *separating* out from the sets that are the values for the variable "s_0," those subsets whose members all satisfy the condition in question. For example, an instance of the axiom schema of separation is

$$\forall s_0 \exists s_1 \forall s_2 \left((s_2 \in s_1) \leftrightarrow \left((s_2 \in s_0) \wedge \forall s_3 \left((s_3 \in s_2) \rightarrow \sim \exists s_4 (s_4 \in s_3) \right) \right) \right)$$

This instance of the schema is an axiom telling us that there is a (possibly empty) subset of every set where the members of that subset are such that their own members do not have any members themselves.

Another indispensable axiom schema shows up in formal theories of arithmetic. This is the axiom schema of induction, telling us that if zero satisfies some condition—that is, the value of the metalinguistic variable "ϕ"—and any number satisfying this condition implies that its successor *also* satisfies the condition, then all numbers satisfy the condition

$$\left(\phi(0) \rightarrow \left(\left(\forall x_0 \left(\phi(x_0) \rightarrow \phi(x_0 + 1) \right) \right) \rightarrow \forall x_0 \left(\phi(x_0) \right) \right) \right)$$

For example, here is one instance of the axiom of induction where the condition specified by the wff instantiating the metalinguistic variable, "ϕ( )," is the condition $\sum_{i=0}^{(\,)} (i) = \frac{i(i+1)}{2}$ (I have omitted the numerical subscripts on the bound variable "x" for readability):

$$\left(\sum_{i=0}^{(0)} (i) = \frac{i(i+1)}{2} \rightarrow \left(\forall x \left(\sum_{i=0}^{(x)} (i) = \frac{i(i+1)}{2} \rightarrow \sum_{i=0}^{(x+1)} (i) = \frac{i(i+1)}{2} \right) \rightarrow \forall x \left(\sum_{i=0}^{(x)} (i) = \frac{i(i+1)}{2} \right) \right) \right)$$

Again, the impossible task of individually specifying every numerical property that we might want to deploy in an axiom of induction is sidestepped thanks to the use of an axiom schema.

Concluding Remarks

Throughout this entry, the notion of a metalinguistic variable has been ubiquitous. Indeed, the distinction between the *object language* (e.g., the formal languages of propositional logic, the predicate calculus, and first-order set theory) and the *metalanguage* (here, the language of English supplemented with symbols like the Greek letters used as variables) is arguably the essential technical distinction underlying the use of axiom schemas. The point of this distinction was made vividly by Tarski (1983, chap. 8) in his proposal of the *T-schema*, an axiom schema for axiomatizing the concept of *true sentence* in a formal language:

For any ϕ, $\ulcorner \phi \urcorner$ is true if and only if S

Here the corner quotes surrounding "ϕ" indicate that we are *not* talking about what it takes for "ϕ" to be true, as this symbol is only a metalinguistic variable over sentences in the object language, not an actual sentence in the object language. Meanwhile, S is a variable representing the *translation* of the value of "ϕ" into the metalanguage. As Tarski helped to make clear, the distinction between the object language and metalanguage is needed to avoid *semantic closure*, in which the language has expressions that refer to other expressions within the same language. As MacFarlane makes clear on pages 61–65 in his textbook, avoiding this kind of closure, and thus avoiding the so-called *semantic paradoxes*, is another primary motivation for making use of axiom schemas in logic and mathematics.

Richard J. Teague

See also Formal Sciences; Logic and Language; Mathematics, 20th Century; Theories, Semantic Conception of; Set Theory

Further Readings

Burgess, J. P. (2022). *Set theory*. Cambridge University Press.

Button, T., & Walsh, S. (2018) *Philosophy and model theory.* Oxford University Press.
Corcoran, J. (2006) Schemata: The concept of schema in the history of logic. *Bulletin of Symbolic Logic, 12*(2): 219–240.
MacFarlane, J. (2021). *Philosophical logic: A contemporary introduction.* Routledge.
Tarski, A. (1983). *Logic, semantics, metamathematics: Papers from 1923 to 1938.* (Edited with introduction and analytic index by J. Corcoran.) Hackett Publishing Company.

Axiomatic Theory

In ancient Greek philosophy, axioms were first principles that required no reasoning or justification as they could be believed to be immediately apparent. In modern philosophy of science, at least in the *non-statement view* of metascientific structuralism, axioms are part of a formal reconstruction of a theory which, according to the founders of this movement in philosophy of science, abated the longish discussion in the community of the translatability between observational and theoretical terms and statements. This entry traces the development of axiomatic theory from its origins, describes the elements of the theory and its models, and concludes with a note on its applications in various sciences.

Observational and Theoretical Statements

Rudolf Carnap, in his *Philosophical Foundations of Physics*, still found "the distinction between what may be called . . . empirical laws and theoretical laws" [. . .] "one of the most important distinctions between two types of laws in science," empirical laws referring to "properties directly perceived by the senses" such as "blue," "hard," "hot" from the point of view of the philosopher, or properties "that can be measured in a relatively simple way" (Carnap, 1966, p. 225). But the distinction between properties perceivable by the senses and those measured in a *relatively* simple way is not really clear-cut, and the simplicity of a measurement is not clearly definable either. Moreover, the problem persists that theoretical laws were formulated in theoretical terms, whereas empirical laws were formulated in observable terms.

Hence, it became necessary to introduce so-called correspondence rules for a translation between the two kinds of terms, for instance between the temperature of a gas as an observable in Carnap's wider sense—temperature being a term of the observational language—and the mean kinetic energy of its molecules as a theoretical term in the theoretical language, or to give an example from the social sciences, between the answer to a question such as "How much confidence do you have in President Joseph Biden?" with possible answers "a lot of confidence, some confidence, not too much confidence, no confidence at all" as an observable and the propensity to cast a vote in favor of President Biden, the latter being a theoretical term, because not even the interviewee would be able to give a numerical value to their probability to vote in the oncoming election for the incumbent candidate. The latter example shows that what is often called the operationalization of a theoretical term can be merely arbitrary with the help of one or more such observable answers to a related question. Another collection of possible answers, for example, 11 or 10 or 5 instead of 4 possible answers, would lead to another operationalization result. To keep this example, one could even argue that one needs some other theory relating the propensity to vote for a certain candidate to the propensity to give certain answers to certain questions, such that correspondence rules would be necessary to link attitudes, voting, and answer propensities to answers given in a certain interview situation.

The so-called syntactic view of theory structure uses these correspondence rules to *translate* between observational and theoretical language (but this kind of *translation* is obviously not of the same kind as a translation between, say, English and Chinese).

The first but incomplete attempt to overcome the dichotomy between observable and theoretical terms was made by Frank P. Ramsey in what is called the *Ramsey sentence of a theory* (also referred to as a *Carnap sentence*), which eliminates the theoretical terms from a theoretical law. If the interpreted theory

$TC = T(\tau_1,\ldots,\tau_n) \wedge C(\tau_{i_1},\ldots,\tau_{i_k};\omega_1,\ldots,\omega_m)$ is the conjunction of all theoretical postulates T (all of them formulated in the theoretical language) and all necessary correspondence rules C (these in turn using terms from both languages, observational and theoretical), then it is possible to replace the theoretical terms τ_{i_j} which occur in the correspondence rules (with its observables $\omega_1 \cdots \omega_n$) with new variables φ_i and replace the interpreted theory with its Ramsey sentence, which postulates that $\exists\varphi_1 \ldots \exists\varphi_k\ TC(\varphi_1, \ldots, \varphi_k, \omega_1,\ldots,\omega_n)$ now claims that substitutions exist for the uninterpreted theoretical terms.

This elimination, however, leaves open the question of how theories are linked together, as in the voting propensity example above. The solution came with the so-called semantic view of theory structure, which introduced set-theoretical entities instead of linguistic ones; this was John D. Sneed's effort to define the logical structure of a theory with set-theoretic predicates instead of speaking about statements whether they are observational or theoretical (which is why this view on theories is called the *non-statement view*).

Elements of the Theory and Its Models

The structuralist program, which builds upon Sneed's work, is centered on the definition of theory elements. A *theory element* is a mathematical structure consisting of a theory core and a set of intended applications, whereby the theory core describes what can be said about the intended applications of the theory core. Theory elements are usually one or a few laws within a scientific discipline, for instance Hooke's spring law (HSL) or classical particle mechanics dealing with masses, forces, velocities, and so on, as the origin of structuralist consideration was clearly within physics, mainly because in physics mathematical formalizations have a long tradition dating at least as far back as Newton and Kepler.

The theory core is a set-theoretic predicate which is typically defined as follows:

> ***T*** is a theory element if and only if there exist sets of partial potential models $\boldsymbol{M}_{pp}$, of potential models $\boldsymbol{M}_p$, of models ***M***, of constraints ***GC***, of links ***GL*** and of intended applications ***I*** such that
>
> $K = \langle M_{pp}, M_p, M, GC, GL \rangle$ is a theory core,
>
> $T = \langle K, I \rangle$ is the theory element and
>
> $I \subseteq M_{pp}$

The set of partial potential models $\boldsymbol{M}_{pp}$ is defining its elements x as another mathematical structure listing all terms which are necessary to speak about the theory in question. In the case of Hooke's law, this list contains a finite set of springs ***S***, a finite set of identical pieces of metal or some other material which can easily brought into an identical form (*weights*) ***W***, three functions $n(w, s)$, $l(w, s)$, and $k(w, s) = l(w, s)/n(w, s)$ yielding the number of weights in a collection w of these weights hanging from a spring and the length of the extension of the spring caused by the weight collection w, whereas $k(w, s)$ is the outcome of an experiment with spring s and weight collection w. In this simple example, the difference between $\boldsymbol{M}_p$ and $\boldsymbol{M}_{pp}$ is very small; it consists of the observation that for small lengths of the extension of the spring, k does not depend on w. The terms s, w, n, l and k can each separately be defined without knowing anything about mechanics: Springs and weights can be identified as such, weights in a collection can be counted, extension lengths can be measured, and the quotient k can be calculated. It is only the assumption that k does not depend on the size of the collection of weights but only on the characteristics of the spring that allows for a redefinition of $k(s) = l(w, s)/n(w, s)$ and separates partial potential models from potential models, as the full model is a potential model for which the axiom of Hooke's spring law holds, namely that for all collections of weights and for each spring which is not extended beyond a certain maximum extension $l_0(s)$, $k(s) = l(w, s)/n(w, s)$ is constant and independent on the collection of weights, from which it follows that $l_0(s)$ and $k(s)$ can be considered as device constants. Moreover, as the experimenter will perhaps use their hands to extend the spring, they will feel Carnap's property directly perceived by the senses as it costs more and more effort the longer the extension is. Hence, the experimenter will assume that there is a monotonically increasing function between their effort and the extension, and it seems reasonably to define this function as a linear function—the HSL-theoretical term *force*—and use it to make this effort quantitatively measurable as proportional to the length of the extension.

The difference between observable terms and theoretical terms is redefined in a straightforward manner as *theoreticity*, always in reference to a theory in question and not an absolute feature of a term. Counting weights, measuring lengths, and dividing numbers necessitate their own theories but not a theory of springs and weights. Hence, the force and the device constants $l_0(s)$ and $k(s)$ are ***HSL***-theoretical terms, all the others are not (and perhaps the value of $l_0(s)$ depends on the accuracy with which $l(w, s)$ and consequently $k(s)$ is measured; the precision of $n(w, s)$ is a lesser problem in this idealized experiment).

Constraints, Links, and Theory Nets

Constraints of a theory collected in the set ***GC*** are about restrictions to its applicability; in the simple example above, this could be the constraint that all pieces are of the same volume, form, and of the same chemical composition (i.e., the theory would not apply to pieces some of which are balls whereas others might be extremely flat but with the same volume) and that these three features are constant over time (i.e., the theory would not apply to identical pieces of ice which might melt away over time).

Links of a theory element collected in the set ***GL*** connect it to other theory elements; in the simple example above, volume and chemical composition are used as if they were properties "that can be measured in a relatively simple way." But to determine the chemical composition of a weight hanging from a spring, some chemistry is necessary; and even to determine the equality of the volume of several metal balls it is necessary to measure their diameters in cm and to assume that the same formula from geometry yielding the volume in cm^3 applies to all of them. Hence, in a way the theoretical links replace the correspondence rules of earlier approaches, but with the important difference that the observational part of a correspondence rule is free from any theory. Links are defined as sets of tuples of terms which connect correspondent terms of the two theories in question. In the example above, there is the term l denoting the length of the extension of the spring in the theory of Hooke springs and the term l^* in geometry, both of which map to integer numbers (multiples of a certain small unit length).

Not only can two theories be linked together, but theory nets can be built that potentially encompass wide areas within and across disciplines. Thus, the discussion about reducing an *observational law*—say the law $pv = kT$ connecting pressure p, temperature T, and volume v of a gas—can be restated in terms of a theoretical law assuming that $p = \frac{2E}{3v} = kT$, where the pressure is related to the mean kinetic energy E of the gas molecules. This also ends the discussion about *reductionism* as the assumption that social science could somehow be *reduced* to psychology, the latter to neurobiology, and further to chemistry, and still further to particle physics; instead theoretical terms of a theory in one of these disciplines could be linked to (instead of reduced to) theoretical terms of another discipline.

In the social sciences, most theories have many more terms than in the simple example from classical mechanics. Nevertheless, a structuralist reconstruction of middle-range theories is possible, too. A theory, call it AOF, about attitude or opinion formation, containing terms such as the propensity to vote for a certain candidate or party and the influence of communicating attitudes between citizens certainly has a number of AOF-theoretical terms—for example, $\alpha : P \times T \to R$ such that $\alpha(p,t)$ is the position of person p at time t on a left–right scale and $\delta : P \times P \to R$— such that $\delta(p_1, p_2)$ and $\delta(p_2, p_1)$ are the movements of Person p_1 and Person p_2 on the left–right scale after they met each other and they have discussed their attitudes toward the candidates and parties running in an oncoming election. In an application of such a theory, the distributions of votes cast for candidates and parties would be observable as AOF-non-theoretical terms (as these could easily be counted) and the distribution of answers to questions like the one cited above ("How much confidence do you have in President

Biden?"). A theory about measuring attitudes with algebraic algorithms such as principal component analysis (PCA) could be linked to AOF, and this intertheoretic link would help to reinterpret the PCA-theoretical terms α and δ as AOF-non-theoretic terms. The function that connects $\delta(p_1, p_2)$ on one hand with $\alpha(p_1, t)$ and $\alpha(p_2, t)$ on the other hand would then be the actual AOF-theoretical term. The theory would have to provide axioms that govern the development of the individual attitude changes and of the distribution of attitudes over a large population in which persons meet each other under certain conditions. Some of these conditions of meeting might be *observable*, in other words AOF-non-theoretical, whereas other such conditions—for instance preferences to discuss with other persons depending on the known or guessed attitude difference between two persons who meet occasionally cannot be directly observed. A few attempts to reconstruct theories of opinion formation have already been made and show the difficulties of such attempts, which are mainly caused by at least two of the big differences between the systems physical theories are about and those which social scientists deal with: The heterogeneity of the *particles* social scientists have to take into account and the fact that human beings, unlike physical particles, can communicate individually and change the internal state of other individuals. Hence, theory nets in economics, and the social sciences generally, but also in the environmental sciences with their extremely complex systems will be much more widespread than in physics. (Balzer, Moulines, and Sneed, the founding fathers of the structuralist program, in their 1987 book themselves reconstructed only pure exchange economics with an intended application that they described as "some specific village at some specific time" (Balzer et al., 1987, p. 161) and did not continue their approach to the formulation of constraints or links.)

Applications in Various Sciences

Since the late 1970s, the method of formally reconstructing theories according to the *non-statement* view has been applied to various (pre-formal) theories in various disciplines. As the movement started within physics, the first candidates to be reconstructed were taken from classical physics, thermodynamics, quantum mechanics, chemistry, biology, medicine, neuroscience, psychology, economics, management, and the social and political sciences. It is not surprising that the density of publications in these disciplines decreases along this list, and it is particularly striking that the acceptance of the *non-statement view* seems to have been greater in German- and Spanish-speaking countries than in English-speaking countries (although English as a publication language still prevails). The disciplines at the end of the list above are much less formalized and quantified than those at the beginning of the list; hence the late and sparse introduction of theory reconstruction into studies of organizations and large social and political systems is understandable, as the reconstruction of a theory that was formulated mathematically long ago is much easier than that of a verbally formulated theory with all its ambiguities. With the emergence and continued development of computational social science, this situation will certainly change in the years to come.

Klaus G. Troitzsch

See also Data Models; Logical Theory, Structuralism

Further Readings

Abreu, C., Lorenzano, P., & Moulines, U. (2013). Bibliography of structuralism III. *Metatheoria—Revista de Filosofía e Historia de la Ciencia, 3*(2), 1–36.

Abreu, C., & Lorenzano, P. (2020). Conversaciones con C. Ulises Moulines. *Metatheoria—Revista de Filosofía e Historia de la Ciencia, 11*(1), 1–31.

Balzer, W., Moulines, C. U., & Sneed, J. D. (1987). *An architectonic for science. The structuralist program.* Reidel.

Carnap, R. (1966). *Philosophical foundations of physics: An introduction to the philosophy of science*. Basic Books.

Nagel, E. (1961). *The structure of science. Problems in the logic of scientific explanation.* Routledge and Kegan Paul.

Sneed, J. D. (1979). *The logical structure of mathematical physics.* Reidel.

Stegmüller, W. (1979). *The structuralist view of theories. A possible analogue of the Bourbaki programme in physical science.* Springer.