Word and Flux

The Discrete and the Continuous In Computation, Philosophy, and Psychology

Volume I
From Pythagoras to the Digital Computer

The Intellectual Roots of Symbolic Artificial Intelligence

with a Summary of Volume II
Continuous Theories of Knowledge

Bruce J. MacLennan

Preface

> Things, then, both those properly so called and those that simply have the name, are some of them unified and continuous, for example, an animal, the universe, a tree, and the like, which are properly and peculiarly called "magnitudes"; others are discontinuous, in a side by side arrangement, and, as it were, in heaps, which are called "multitudes," a flock, for instance, a people, a heap, a chorus, and the like. Wisdom, then, must be considered to be the knowledge of these two forms.
>
> — Nicomachus of Gerasa (c. 60 – c. 120 CE)

We live at the end of an era. For 2500 years philosophy and science have been dominated by the view that knowledge can be represented in discrete symbol structures and that thought is the formal manipulation of these structures. This view has been so pervasive that it has rarely been recognized as an assumption worthy of criticism. Its dominance has been strengthened by the fact that the occasional criticisms of it have seemed too indefinite to merit scientific consideration. Nevertheless, as a consequence of its wide acceptance, the implications of the traditional view have been extensively explored, so that now its weaknesses have become apparent.

Crucial to this process has been a technological development: the digital computer, for the digital computer is the formal symbol manipulator *par excellence*. Therefore the availability of high-speed computers has permitted the *empirical* investigation of the traditional view of knowledge; in particular, the modest success, until recently, of artificial intelligence has exposed the limitations of the traditional view.

Throughout its 2500 year history the tradition has had its dissenters, but there are two reasons that it has not had significant competition until now. First, the tradition had to be worked through to its conclusion, which has culminated in the theory of computation and "symbolic" artificial intelligence. Second, the alternative views seemed inherently nonscientific, and so they have been generally unacceptable to our increasingly scientific culture. This situation changed abruptly in the 1980s with the emergence of *connectionism* and *neural network theory*, which provide the basis for a scientific account of an alternative theory of knowledge. Thus two related events, the final working out of the traditional theory and the emergence of a scientific alternative, have combined to precipitate a revolution in the theory of knowledge. It is perhaps the most significant change in our understanding of the fabric of knowledge in two and a half millennia; we indeed live at a historic moment.

This book and its intended companion are simply the detailed presentation and justification of the preceding claims. They originated in a graduate course, "Epistemology for Computer Scientists," which I taught in the mid to late 1980s to address a general lack of up-to-date philosophical knowledge among AI researchers. In the early 1990s this developed into courses and seminars intended to explain the radical reorganization of epistemology implied by connectionism and artificial neural network approaches to cognitive science and AI.

The title of the two volumes, *Word and Flux*, derives from a fundamental tension between discrete and continuous phenomena that has permeated our culture's view of knowledge from its earliest attempts at philosophical thinking. This first volume is divided in two parts. Part I traces the traditional view of knowledge from its origin in ancient Pythagoreanism, where we first find the attempt to reduce continua to discrete symbol structures — to *arithmetize geometry*. The issue of this attempt includes formal logic and the notion that thought is (digital) computation. Part II investigates the acceleration of this process in the nineteenth and twentieth centuries, including the apparently successful arithmetization of geometry, the attempts to formalize mathematics and science, and the computational theories of mind that have dominated cognitive science and artificial intelligence. This acceleration also brought with it the first signs of weakness in the traditional view through the discovery, in the first half of the twentieth century, of the theoretical limitations of discrete symbol systems.

Volume II, which was never completed, intended to explore the history of an alternative view of knowledge.[1] The criticisms that have been made against the tradition over the centuries provide the insight necessary to see both the weaknesses of the tradition and the requirements for an alternative. Volume II was also intended to include a systematic presentation of the alternative — at least as systematic as was possible at that early date. The foundation is provided by the theory of artificial neural networks and massively parallel analog computation. Volume II would have concluded with a

[1] See Chapter 11 for a detailed outline of the unwritten chapters of Volume II.

discussion of the implications of this theory for our understanding of knowledge, in general, and for our understanding of the mind and of science, in particular. I regret that other research and writing activities distracted me from completing Volume II, but perhaps the completed part of Chapter 10 and the detailed outline in Chapter 11, which together constitute Part III of this book, will compensate to some degree. *Ars longa, vita brevis!*

Bruce MacLennan
Douglas Lake
Tennessee

Contents

Part I

The Archaeology of Computation

Chapter 1

Overview

1.1 Method of Presentation

The Psychoanalysis of Computation

Connectionism challenges many of the assumptions of artificial intelligence, cognitive science and epistemology. Unfortunately, like most assumptions, these reside in a "background" of which we're hardly aware. Therefore, to better understand the significance of connectionism we must first bring these assumptions out into the light.

Our goal will be a kind of "psychoanalysis" of the traditional theories of cognition and computation. As in psychoanalysis, we will expose long repressed beliefs and desires. The object is not necessarily to eliminate these beliefs and desires, but rather, by bringing them into conscious awareness, to open them up for critical scrutiny.

The Archaeology of Computation

Another metaphor for our task is an archaeological investigation, which in fact it is, since *archaeology* means *the science of beginnings*. Thus, just as the hill of Troy has been peeled back, stratum by stratum, to reveal its history, so we may dig below the edifices of the cognitive and computational sciences so that we may understand their historical supports (Fig. 1.1).

Advantage of Historical Presentation

Although the psychoanalytical and archaeological metaphors may help us understand our task, it is in fact better to present the material in its historical order. It's easier to trace from the root out to the leaves, than from the leaves back to the root. The reason is that when we begin with the leaves there are too many branches to be followed simultaneously, and their ultimate destination is unclear. On the other hand, if we begin at the root then we'll always be following just one branch at a time. When we want to

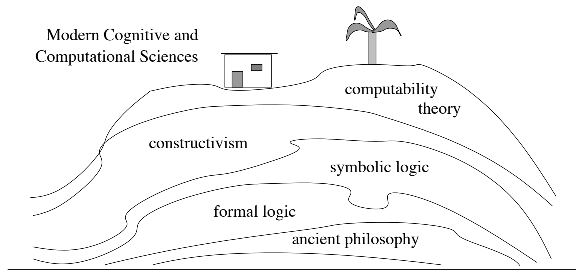


Figure 1.1: Archaeology of Computation

Additional Information

Occasionally you will find supplementary text, such as this, separated from the main text by horizontal rules. This supplementary text takes many forms, including definitions, explanations, digressions, technical details, amplifications, and pointers into the literature. You may read this material or not, as your interests, background and time permit.

---

explore another branch we will go back to the branch from which it grows, and this will be a branch with which we're already familiar. The historical approach, by mimicking the genetic development of the ideas, keeps their relationships clear.

Disadvantage of Historical Presentation

On the other hand, the historical approach has a disadvantage. Since the roots of the modern cognitive and computational sciences go very deep indeed, it will not be obvious how these roots are connected to the issues central to these sciences. Of course, I could simply assert that they are, and that you will see the relevance when we finally get to the leaves, but that's not a very satisfactory approach. Instead I'll very briefly go from the top down, showing you the connections between the cognitive and computational sciences and some perennial philosophical and scientific concerns. This will give you a kind of prophetic insight when we then turn around and follow the historical process.

1.2 Artificial Intelligence and Cognitive Science

Artificial Intelligence and Natural Intelligence

Artificial intelligence (AI) began about 1960 as the attempt to get computers to act with human-like intelligence, although from the beginning many AI researchers were also interested in natural intelligence. It has always seemed likely that understanding the processes of natural intelligence would facilitate their emulation in our artifacts, and conversely that the ability to create intelligent artifacts would inform our theories of natural intelligence. As a result developments in artificial intelligence and cognitive science have proceeded in parallel, with many fruitful exchanges of ideas.

Computation and Calculi

Central to both artificial intelligence and cognitive science have been the use of logic-based knowledge representation languages and rule-based descriptions of cognitive processes, which are based on the idea of a *calculus*, that is, a set of symbolic structures and rules for the mechanical manipulation of these structures. This is the core idea of computer science, since the digital computer is nothing more than a mechanical rule manipulator, a physical device for implementing calculi. Although the idea of a calculus is very old — we will see that it comes from ancient Greek philosophy — it was not well understood until the theory of computation was developed in the first decades of the twentieth century.

1.3 The Theory of Computation

Theory of Computation Predates Computers

Surprisingly, the *theory* of computation predates by a decade or more the *invention* of the first electronic digital computers in the 1940s. Why were mathematicians and logicians in the 1930s worried about the theory of computation when there weren't any computers? To understand this we must investigate some of the issues that concerned mathematicians and philosophers in the late nineteenth and early twentieth centuries.

Cantor's Set Theory

In the late nineteenth century Georg Cantor developed a remarkable theory of sets that included a series of infinite numbers of greater and greater magnitude.[1] As we'll see, mathematicians and philosophers have always been uncomfortable with the infinite (both the infinitely large and the infinites-

[1] To avoid clutter, I am omitting the references to the literature, which will be found in the later chapters where these topics are discussed in detail.

mal, or infinitely small); this attitude is rooted in ancient philosophy. As a result, many mathematicians doubted whether Cantor's theory was meaningful. What did it mean to claim that these infinite sets and numbers exist? Nevertheless, Cantor's powerful methods were eventually widely adopted by mathematicians, and by now have become an accepted part of mainstream mathematics.

Constructivism

There were — and are — many mathematicians who do not accept the proof techniques used by Cantor and most other mathematicians. One of the first of these was L. E. J. Brouwer (1882–1966), who founded a school of mathematics known as *intuitionism*. His ideas were further developed by a group of mathematicians who are known as *constructivists*, because they believe that it makes sense to talk about the existence of a mathematical object only if it can, at least in principle, be constructed. Thus, a constructivist would not accept a proof of the existence of a number satisfying an equation, unless that proof showed how to construct (i.e., calculate) that number. The goal of the constructivists was to prove as much as possible constructively; what could not be so proved was to be discarded as chicken scratches on paper.

Constructivism and Computation

What do these disputes in the philosophy of mathematics have to do with computers? The constructivists had a problem. Before they could offer a definite alternative to nonconstructive mathematics they would have to be able to say precisely what they meant by a number (or any other mathematical object) being constructible. To this end they developed the *theory of effective calculability*, or, as it's now more commonly called, *computability theory*. Roughly, something is *effectively calculable* or *computable* if it can be calculated by the strictly mechanical application of formal rules in a finite number of steps and using finite resources. Specifically, it is *effectively calculable* just when there's a *calculus* for *calculating* it. It is no coincidence that all these words contain the root *calcul-*; the exact significance is discussed in detail later (p. 26).

Thus, the efforts of these mathematicians and logicians led to the theory of computation, which is the core theory of computer science, and provides many of the techniques and central insights of traditional AI and cognitive science.

The Preference for Calculi

We have traced briefly the cognitive and computational sciences back to the emergence of constructivist ideas in the beginning of this century. These ideas made precise the notion of the mechanical manipulation of symbolic structures. But the roots of the preference for calculi — the preference for rules that can be applied mechanically and for the representation of knowledge in symbolic structures — go much deeper, and manifest themselves in several themes that run through Western philosophy's view of knowledge.

1.4 Themes in the History of Knowledge

Formalization of Knowledge

The first theme is that true, scientific *knowledge* can be formalized, that is, expressed in finite formulas constructed from discrete symbols (such as the letters, numbers and other signs of mathematical formulas). A complementary part of this theme is that scientific *inference* can be expressed in a finite number of precise, definite rules for manipulating these formulas, and that these rules depend only on the shape and arrangement of the symbols (think of algebraic processes such as solving a simple equation). This theme began with Aristotle's identification of formal laws of deduction, continued with the extensive logical investigations of the Middle Ages (by Ockham et al.), showed its potential through attempts by Lull, Leibniz and others to devise a universal calculus for knowledge representation and inference, and reached maturity in the nineteenth century in the symbolic logics developed by philosophers and mathematicians such as Boole, Frege, and Russell.

Logical Positivism

In the first decades of the twentieth century the movement to formalize scientific knowledge developed into an extreme form called *logical positivism* (or *logical empiricism*). This philosophy of science is based on Russell's logic and Wittgenstein's theory of *logical atomism*, which views items of scientific knowledge as logical constructions of elementary propositions, such as sense data or measurements. The philosophy was explicitly formulated in the "Vienna Circle," whose members included Moritz Schlick, Rudolf Carnap and Otto Neurath, but rapidly spread to become the dominant philosophy of science in the first half of the twentieth century. Many major movements in the individual sciences, such as the Copenhagen interpretation in quantum mechanics (Bohr), functionalism in sociology (Durkheim) and behaviorism in psychology (Skinner), are simply applications of logical positivism to that science. Indeed, although it began to lose popularity about 1960, logical positivism still forms much of the background of our view of "real science." In particular, although contemporary cognitive science has rejected behaviorism, a positivist view of *method*, it has not completely broken away from rule-based knowledge representation, a positivist view of *cognition*.

Rationalization of the Continuum

The second theme is usually called the *arithmetization of geometry*, but is more accurately called the *rationalization of the continuum*. It stems from the discovery by the ancient Pythagoreans that certain continuous magnitudes (what we now call *irrational numbers*) could not be expressed as ratios of whole numbers; that is they could not be *rationalized*, or expressed as finite, discrete formulas. Since the ancient Greeks assumed that true knowledge was by its nature expressible in finite, discrete formulas, there seemed to be something inherently irrational (incomprehensible) about continuous magnitudes. Thus arose the problem of *arithmetizing geometry*, that is, of reducing the troubling continuous magnitudes, studied in geometry, to the more comfortable discrete numbers (integers), studied in arithmetic. The ancient Greeks were not able to accomplish this — Euclid axiomatizes arithmetic and geometry separately — and the problem of arithmetizing geometry remained open until the nineteenth century when Dedekind and Wierstrass showed how real numbers could be constructed from sequences of rational numbers.

Limitations of Calculi

These accomplishments meshed into the development of formal logic, so that by the first decades of the twentieth century it seemed that mathematical knowledge and reasoning would be soon reduced to a calculus. Unfortunately, just when it seemed that the rationalization of mathematics was at hand, disturbing and seemingly irreparable defects were discovered in the foundations; mathematical logicians such as Skolem, Löwenheim, Gödel and Turing discovered fundamental limitations in the ability of calculi to represent knowledge and inference. In particular they call into question the apparent success of the arithmetization of geometry. Although most of these discoveries were made before 1940, their implications for science and mathematics and for cognition and computation are still unclear and a topic of continuing debate. Part of the difficulty is that the assumptions of discrete knowledge representation are so embedded in our view of knowledge that we find it difficult to see any alternative to viewing knowledge as a calculus.

Limitations of Symbolic AI

The development of electronic digital computers in the 1940s provided the first opportunity to directly test the hypothesis that knowledge was representable by finite, discrete formulas and that it could be processed by discrete symbol-manipulation operations. Early experiments were very promising, and led in the early 1970s to a surge of interest in rule-based systems. Unfortunately, more ambitious goals in both artificial intelligence and cognitive science began to expose the weaknesses of the calculus as a model of cognition; these include the inability to handle novelty, minor exceptions, uncertainty and context-sensitivity without an exponential explosion in the number or size of the rules, and thus in the required computational resources. Further evidence of the inadequacy of the calculus model comes from the "cognitive inversion" of rule-based approaches: what they seem to do well are higher cognitive activities (formal logic, chess, mathematics); what they seem to do poorly are lower cognitive activities (perception, coordination, pattern recognition, image processing). This inversion suggests that natural intelligence is not rule-based. Thus, by the early 1980s researchers were being driven to the conclusion that the calculus was impractical as an implementation of artificial intelligence and implausible as a model of biological intelligence. But what could be an alternative?

1.5 Alternative Views of Cognition

The Traditional View: A Calculus

An alternative is *connectionism*, but to see its importance it is necessary to understand it in the context of alternative theories of knowledge. We have seen that the dominant view of knowledge in the Western philosophical tradition has reflected a pervasive aversion to the continuous and the infinite, and a preference for definite rules and finite, discrete, symbolic structures. Thus the calculus becomes the ideal model of a knowledge structure, and the goal of any science becomes the axiomatization of its knowledge. Further, since the essence of a calculus is in its formal structure rather than its material embodiment, there has been a tendency to think of knowledge as independent of its biological context.

Philosophical Alternatives

There have always been philosophers that have disagreed with this view of knowledge. For example, in ancient Greece Heraclitus stressed the continuity of knowledge, both in its structure and its evolution; the Sophists discussed the dynamics of knowledge and its relation to society; and Aristotle investigated analogical knowledge structures and the role of knowledge in its biological context. Positivist theories of knowledge were also criticized in the eighteenth and nineteenth centuries by philosophers such as Kant, Hegel, Dilthey and Nietzsche. The early twentieth century brought the development of two important alternatives to positivism, *phenomenology*, especially as formulated by Husserl and Heidegger, and *ordinary language philosophy*, as investigated by Austin, Wittgenstein and others. The phenomenologists stressed the importance of everyday understanding, including the social context, as a foundation for all knowledge; the ordinary language philosophers showed the impossibility of defining the terms of everyday language in the way demanded by the logical positivists, and the inadequacy of a positivist model of conceptual knowledge. The latter half of the twentieth century has also seen a renewed attempt to understand knowledge in both its biological (Lakoff, Johnson, Millikan, et al.) and social contexts (Kuhn, Lakatos, Popper, Dreyfus, et al.).

Psychological Alternatives

There have also been minority views in psychology, most notably *Gestalt psychology* (Wertheimer, Koffka, Köhler, et al.), which in the early twentieth century opposed a continuous, holistic view of knowledge to the discrete, atomistic models of behaviorism. However, before it was able to develop an empirically verified theory of the mechanisms of cognition, it fell before the logical positivist juggernaut. Nevertheless, the decline of positivism has brought a renewed interest in Gestalt ideas, as well as investigations into mental imagery (Shepard, Miller, Paivio, Kosslyn, et al.) and the inadequacy of discrete, symbolic models of concepts and categories (Rosch et al.). There has also been a trend toward putting cognition in its biological context (Gibson, Wason, Johnson-Laird, et al.).

Neuroscientific Alternatives

By 1950 the view that the brain stores information like a computer memory, with discrete items of information stored in discrete locations, had come under attack from neuroscientists such as Karl Lashley. He proposed that each "item" of information was distributed over an entire region of the brain, with many different items being superimposed in the same region. Although Lashley was unable to explain the mechanism of distributed memory, such a mechanism was proposed in the 1960s by Karl Pribram, who proposed that memory operated on principles analogous to an optical hologram, that is, the interference of continuous fields. This idea forms a bridge between the theories of the Gestalt psychologists and modern connectionist theories.

1.6 Connectionism

Origins of Connectionism

In the middle decades of this century, before the sequential, digital computer had completely dominated our view of computation, there were many exciting applications of *artificial neural networks*, parallel analog computer systems with principles of operation patterned after the brain, such as Rosenblatt's "perceptrons" and Selfridge's "Pandemonium." For a variety of reasons, both philosophical — the unquestioned acceptance of logical positivism working its way through computer science — and technological — the rapid progress in digital electronic technology — most neural network research had been abandoned by the late 1960s. Although a few researchers continued in the area (e.g., Grossberg, Longuet-Higgens, Anderson, Willshaw), the 1970s were the decade of rule-based models of intelligence, whether natural or artificial.

What Computers Can't Do

Hubert Dreyfus' 1972 book, *What Computers Can't Do*, is best known for its apparent negative conclusion, the impossibility of artificial intelligence, but it is also important for opening the way for connectionism since, in addition to showing the limitations of traditional AI, it reviewed some of the psychological evidence against symbolic knowledge representation, showed the importance of the phenomenological view of knowledge, and pointed to holographic models of cognition as an alternative to rule-based computation.

Contemporary Connectionism

By the early 1980s growing dissatisfaction with the calculus model was causing some researchers in cognitive science and artificial intelligence to explore alternative models based on the cooperative and competitive interaction of large numbers of simple elements that represent and process information at the subsymbolic level (Feldman, Ballard, Rumelhart, McClelland, Hinton, Smolensky, Sejnowski, Sutton, Barto, Kohonen, Hopfield, Hofstadter, and many others). Although some of the earlier connectionist models still used discrete (digital) symbols stored locally in a network, the trend has been toward representations based on continuous (analog) values distributed throughout the network. The result is both psychologically and biologically more plausible. Psychologically, because connectionist systems exhibit many of the same strengths and weaknesses as natural systems; for instance, they do not exhibit the cognitive inversion. Biologically, because it is much easier to conceive of an implementation in nervous tissue of a connectionist system than of a rule-based system. Since the mid-1980s there has been an efflorescence of connectionist research, both basic and applied, but its implications for our understanding of knowledge, language and cognition – both natural and artificial — are still on the horizon.[2]

Significance of Connectionism

The contribution of connectionism is that it integrates these alternative theories of knowledge into the body of scientific theory by providing access

[2] The terms *connectionism*, *parallel distributed processing* (PDP), and *neural networks* are used nearly synonymously. I prefer *connectionism* to the more awkward (but perhaps more accurate) *parallel distributed processing*, but both refer to the same alternative theory of knowledge and cognition. I will reserve the term *neural networks* (and *neurocomputing*) for the new approach to computing inspired by biological neural systems. The difference, however, is more of intent than of content.

to the mechanisms of cognition. That is, although many philosophers and psychologists were aware of the inadequacy of positivist theories of knowledge, their alternatives seemed unscientific because they did not show how they related to the commonly accepted theories and methods of the other sciences.

The Homunculus

The problem is to "banish the homunculus" — a colorful description of the goal of a scientific psychology. The problem is shown quite literally in some popular presentations of the brain. We may see the eyes and other senses wired to dials and TV monitors within a control room in the brain. The muscles, internal organs, etc. are in turn controlled by wires connected to control levers and switches in this same control room. And there, in the middle of it all — looking at the dials and manipulating the levers — is a little man (homunculus), usually wearing an engineer's cap!

It is apparent that nothing has been accomplished by explaining our cognitive capabilities in terms of a little man in our heads. How are his cognitive capabilities to be explained? By reduction to another, smaller homunculus?

Although few scientists would postulate a homunculus in such an explicit form, it is remarkably hard to keep homunculi out of theories of the mind. For example, if a neuroscientist identifies a visual projection area in the cortex, the question immediately arises: Who (or what) is "looking" at the image in that area? No one proposes tiny TV screens in our brains, but a visual projection area is not so different unless we can explain mechanistically how the image is processed.

Banishing the Homunculus

Thus the only way to be sure we have banished the homunculus is to give a complete in-principle mechanistic (i.e., non-homuncular) account of every step from sensation to action; only then can we be sure we haven't begged the central question.

Part of the importance of AI to cognitive science is that it promises to do exactly this. Since, we suppose, there is no "ghost in the machine," successful artificial intelligence is ipso facto a non-homuncular theory of intelligence. Unfortunately, traditional AI was not successful, since it was based on an inadequate theory of knowledge. The promise of the new AI is that, by virtue of a better epistemology, it will succeed in banishing the homunculus.

Chapter 2

The Continuous and the Discrete

2.1 Word Magic

> All words are spiritual, nothing is more spiritual than words.
>
> — Walt Whitman

> He who shall duly consider these matters will find that there is a certain bewitchery or fascination in words, which makes them operate with a force beyond what we can naturally give account of.
>
> — Robert South

We can never be wholly free of our background of assumptions, but we can become more aware of it, and in this way expose it to change. Although I claimed in the last chapter, and will show in detail in this chapter, that the traditional theory of knowledge grows out of attitudes prevalent in early Greek philosophy, in fact they are grounded in a reverence and awe of language that is common to all cultures.

For if we look, especially in less scientific cultures, we find magic power attributed to words.[1] It's well known that in many societies everyone has a

[1] Sources for this section are Cornford (*FRP*), Englefield (*Lang.*, Ch. 11), Frazer (*GB*, pp. 244–262), Frazer (*NGB*, pp. 235–246), Ogden & Richards (*MoM*, Ch. 2).

secret name that's known only to one's closest family, because it's believed that anyone who knows an individual's true name (the secret name) has power over that individual. Indeed, a person's name and soul are effectively identical. Likewise, many religions believe that knowing the name of a god grants some control over that god, or that the name of a god should not be spoken out loud (hence in Judaism God's "unspeakable name," represented by the tetragrammaton 'YHWH' יהוה, is pronounced *Adonai*, "Lord").

This magical power is not limited to personal names, for, as Cornford (*FRP*, p. 141) says,

> To classify things is to name them, and the name of a thing, or of a group of things, is its soul; to know their name is to have power over their soul.

You may wonder how words came to be invested with such power. One theory is that the earliest forms of communication were imperative, and that many magical procedures had their origin in verbal and nonverbal commands: spoken orders (spells), gestures, facial expressions (the "evil eye"), pantomime (ritual dances), etc. (Englefield, *Lang.*, pp. 124–127). However, in the case of word magic a more direct source is apparent, for in many cases words do in fact operate directly to produce an effect. Verbal formulas of this kind are called *performatives* (Austin, *PP*, Ch. 10) because they perform some action. A familiar example of a performative is the formula "I now pronounce you husband and wife." The mere uttering of this phrase by an authorized person (legally or religiously ordained) in an appropriate ceremony is sufficient to make the marriage a *fact*. This is true in general: performatives do not ask or even command that something be done; they *do it*.

Performatives often begin with formulas such as "I hereby ..." or "By the power vested in me ..." that signal the special nature of the utterance. They have causal efficacy only if uttered in the appropriate circumstances (e.g., a marriage, graduation or other ceremony) by someone duly authorized. Some performatives require no special authority, such as "I apologize" or "I promise," but even in these cases they may not be efficacious if uttered by a young child, by a mentally incompetent adult, or under duress, etc.

The connection with word magic should be clear. In all societies, but especially in authoritarian ones, many states of affairs can be created by an authorized individual uttering the appropriate verbal formula. Marriage, banishment, official office, death, kinship, identity, possession, access to food or shelter — all may be granted or refused by speaking the right words in the right way. Is it any wonder that the power came to be attributed to the words themselves rather than to the social context of their use?

> Language, that stupendous product of the collective mind, is a duplicate, a shadow-soul, of the whole structure of reality; it is the most effective and comprehensive tool of human power, for nothing, whether human or superhuman, is beyond its reach. (Cornford, *FRP*, p. 141)

Thus it is hardly surprising that we should find in the earliest philosophy an attempt to capture the world by verbal formulas.

2.2 Pythagoras: Rationality & the Limited

> What is the wisest thing? Number; but second, the man who assigned names to things.
>
> — Pythagoras (attributed in Iamblichus, *Vita Pythagorae* 82; DK 58C4)

> [The Pythagoreans] took numbers to be the whole of reality, the elements of numbers to be the elements of all existing things, and the whole heaven to be a musical scale and a number.
>
> — Aristotle, *Metaphysics* 1.5.985b23 (DK 58B4)

> And indeed all the things that are known have number; for it is not possible for anything to be thought of or known without this.
>
> — Philolaus (DK 44B4)

> There is divinity in odd numbers, either in nativity, chance or death.
>
> — Shakespeare, *The Merry Wives of Windsor*, 5.1.2

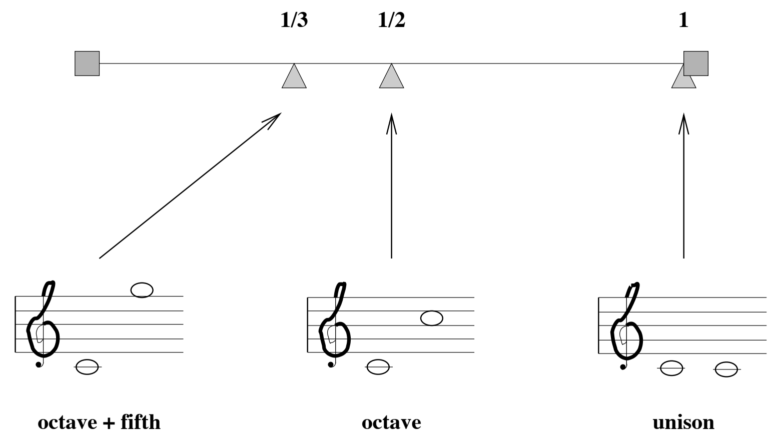


Figure 2.1: Intervals and ratios of lengths. If the string is divided in half (and all other factors are kept constant), then the string sounds an octave higher. If the string is divided in thirds, then it sounds an octave and a fifth higher.

2.2.1 Discovery of the Musical Scale

Pythagoras, 572–497 BCE

Discovery of the Musical Scale

The early Pythagoreans — perhaps Pythagoras himself — discovered the relationship between musical intervals and ratios.[2] They discovered that strings divided in the ratio 1:2 sounded consonant, producing the interval that we call an octave. Similarly, a ratio of 2:3 produces the interval of a perfect fifth, and 3:4 produces a perfect fourth (Figs. 2.1 and 2.2). It might seem that this discovery's main significance is in music, but in fact it became a paradigm for most later science, logic, mathematics and philosophy. This claim will take some justification, and that is the aim of this section. To understand the significance of this discovery, it's important to observe that tuning a musical instrument is a skill that requires some training and expertise. It is not easy to describe how the instrument sounds when it's in tune. Rather, the teacher must show the students, who must learn to recognize the difference with their own ears. In this sense, tuning is apparently inexplicable; that is,

[2] Much of this discussion of the Pythagoreans is drawn from Burnet (*GPI*, §§ 28–38 and §§ 69–75). Another good source is Kirk, Raven & Schofield (*Presoc.*), although the first edition (Kirk & Raven, 1957) is more outspoken in its opinions. Greek philosophical terms are discussed in Peters (*GPT*). I have also used Liddell, Scott & Jones (*LSJ*), and occasionally Donnegan (*Lex.*). Other sources for the Pythagoreans are Burkert (*LSAP*), Sinnige (*M&I*) and Maziarz & Greenwood (*GMP*).

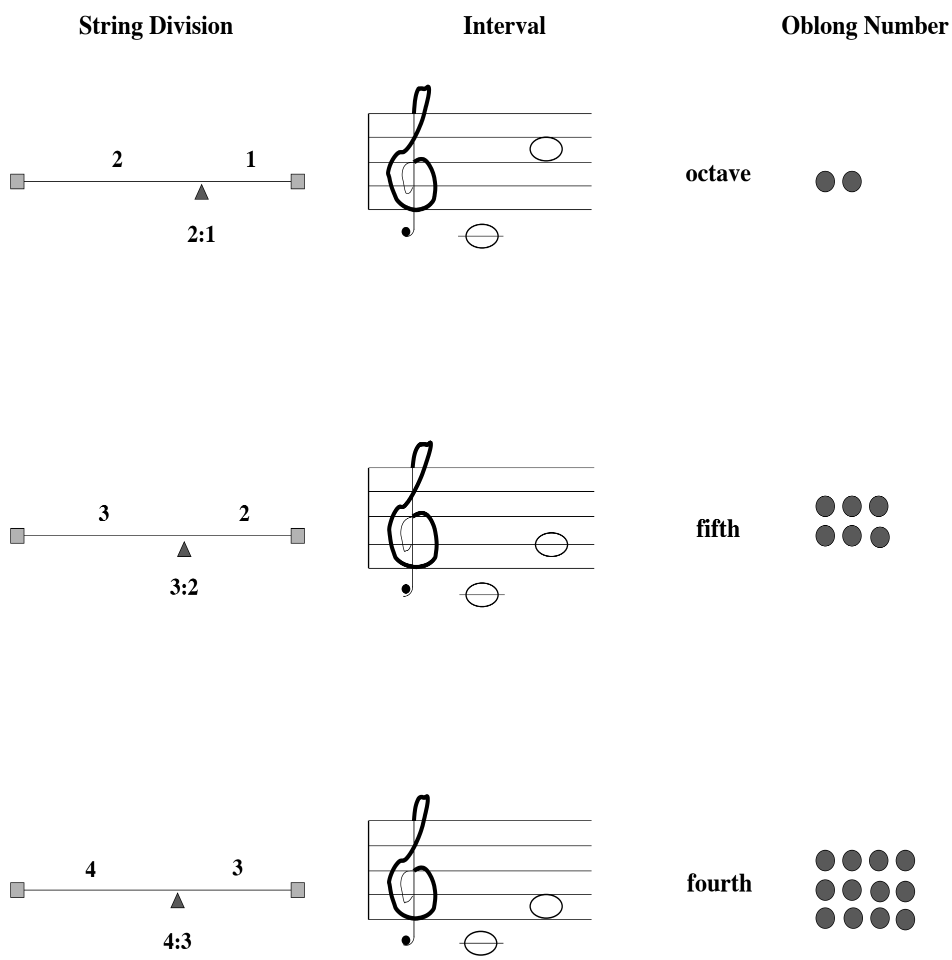


Figure 2.2: Musical intervals based on ratios. If the string is divided in the ration 1:1, then its halves sound the same pitch, called *perfect unison*. The "oblong numbers" (ratios of the form $n + 1 : n$) determine progessively more dissonant intervals. If the string is divided in the ratio 2:1, then its parts sound in the interval of an *octave*, which is the most consonant interval after unison. The ratio 3:2 produces a *perfect fifth*; 4:3 a perfect fourth; 5:4 a major third, 6:5 a minor third; and so on.

we cannot explain it in words.

Reduction of Expertise to Ratios

The accomplishment of Pythagoras was to show that tuning *is* explicable. Specifically, he showed that being in tune is equivalent to satisfying certain ratios. The measurement of these ratios, in turn, is a simple procedure that does not require a "well-trained ear." As Maziarz & Greenwood (*GMP*, p. 43) say,

> Intervals between sounds perceptible only to the fine ears of expert musicians, which could be neither explained to others nor referred to definite causes, were now reduced to clear and fixed numerical relations.

The impact of this discovery on Greek thought was profound. Burnet (*GPI*, p. 56) claims that the concordant intervals

> yield the conception of 'form' as correlative to 'matter', and the form is always in some sense a Mean. This is the central doctrine of all Greek philosophy to the end, and it is not too much to say that it is henceforth dominated by the idea of [harmonia] or the tuning of a string.

(Note that Greek *harmonia* (ἁρμονία) doesn't mean harmony in the modern sense: "the word 'harmony' ... means in the Greek language, first, 'tuning,' and then 'scale'" (Burnet, *GPI*, p. 45).)

In modern terms, what the Pythagoreans accomplished was to reduce a kind of *expertise* (tuning) to a simple *rule* (a ratio). Thus it is both an example of the reduction of an expert judgement to computation, and an example of embodying a phenomenon of nature in a mathematical law. Next we'll discover why it was of crucial importance to the Greeks that the rules was expressed as a ratio.

2.2.2 The Rational

Etymological Considerations

Occasionally, but especially in this chapter, we will consider the origin of some word or group of words. We make these etymological forays for several reasons. First, the histories of these words are part of the archaeology of the theory of knowledge; they exhibit ancient habits of thought from which we derive our own habits. Second, since this book is concerned with knowledge representation, and especially with the role in it of concepts and language (recall its title, *Word and Flux*), therefore these historical data become examples of the very phenomena of interest. They show us the complexity, in actual use, of the meaning of certain key words, and how the constellation of meanings of such a word may influence the ways we think about the world. Thus these etymological discussions should be read both as pertaining to the history of the theory of knowledge and as illustrating the interplay of language and cognition.

Logos *in Ancient Greece*

We begin by considering the way the Greek word *logos* ($\lambda\acute{o}\gamma o\varsigma$) was used in Pythagoras' time, which will help us appreciate the significance to the ancient Greeks of the reduction of a natural phenomenon to ratios. This word ultimately derives from the verb *to say* (*legō*), and so the most basic meanings of *logos* relate to saying. By the time of Pythagoras, *logos* could mean word, language, talk and thought. In a more extended sense it could refer to verbal accounts of things, such as reasons, explanations, principles, meanings and causes. Finally, *logos* could mean a ratio or calculation, which is an explanation or reason in the mathematical domain. From *logos* we of course get such terms as *logic* and *logical* as well as the *-ology* that ends the names of many sciences.

It is important to realize that for the ancient Greeks these meanings formed a whole. Thus, that which had a *logos* was simultaneously that which was reasonable, explainable, principled, meaningful, reducible to causes, thinkable and sayable. Conversely, *alogos* came to mean irrational, inexplicable, unprincipled, meaningless, causeless, incomprehensible and unspeakable. (See also p. 31.)

It's easy to see how the Greeks would view the discovery of the musical scale as a triumph of reason over unintelligibility, as indeed it was. The Pythagoreans believed that just as tuning had been reduced to ratios, so eventually all phenomena would be reduced. Hence their claim: "Everything is number."[3]

The Rational and Ratios

We need to make one more linguistic observation. The Latin word *ratio* was used with a similar constellation of meanings to the Greek *logos* (in part

[3]There is disagreement about whether the Pythagoreans said everything *is* number or everything *is like* number; indeed their position may have changed over time. Aristotle states quite clearly however that the Pythagoreans thought numbers were the actual material constituents of things, and that in this they differed from the Platonists (Aris., *Met.* 1.6.6.987b28–29; see also quotation on p. 17). It is noteworthy that *harmony* and *arithmetic* both derive from the same Indo-European root *ar-*, meaning 'to join together' (*AHD*).

because it was used to translate *logos*). In extended senses *ratio* meant a reckoning, account, computation, calculation, list, catalog, relation with or reference to, plan or procedure, principle, reason, ground, method, order, rule, theory, system, knowledge, opinion, or ratio. *Ratio* is of course the source of our word *rational* and its derivatives. It is thus no coincidence that in English *rational* can mean both *expressible as a ratio* and *intelligible*. To the ancients, what was intelligible was what was expressible in words, and a numerical ratio was the paragon of such expressions. Thus, to the Pythagoreans, the rational — in the sense of intelligible — was identical with the rational — in the sense of reducible to ratios.

The connection between what we may call *mathematical* rationality and *epistemological* rationality may seem no more than a historical curiosity, but we will see that over the centuries the two notions have influenced each other in mathematics, logic, philosophy and computer science.

Pythagorean Number Theory

We turn next to the Pythagorean theory of numbers. This will help us understand their idea of ratio (logos). More importantly, however, we will see that it is the ultimate root of formal logic, some critical issues in the foundations of mathematics, the theory of computability, and knowledge representation languages.

The Pythagoreans represented numbers by pebbles, for example, •, ••, •• •. By placing these pebbles in various arrangements they were able to demonstrate (but not prove in the modern sense) a number of elementary properties of numbers. For example, the triangular numbers can be arranged into an equilateral triangle, which shows that each triangular number is the sum of consecutive integers (Fig. 2.3):

$$\begin{aligned} 1 &= 1, \\ 3 &= 1+2, \\ 6 &= 1+2+3, \\ 10 &= 1+2+3+4, \\ &\text{etc.} \end{aligned}$$

(The Pythagoreans also recognized square, pentagonal, hexagonal numbers, etc.[4]) Similarly, the Pythagoreans were able to prove that the square numbers are the sums of consecutive odd numbers (Fig. 2.4). Notice that if the shape of a *gnomon* (carpenter's square, or rule) is drawn in the figures, then the

[4] See Nicomachus, *Intro. to Arith.* 2.8–12 (Cohen & Drabkin, *SBGS*, pp. 7–9).

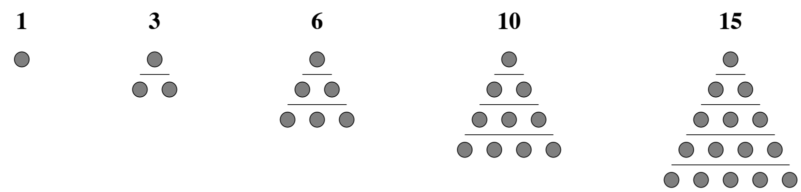


Figure 2.3: Triangular numbers. Applying the "rule" shows that consecutive triangular numbers are the partial sums of the natural numbers, 1, $1+2$, $1+2+3$, $1+2+3+4$, ...

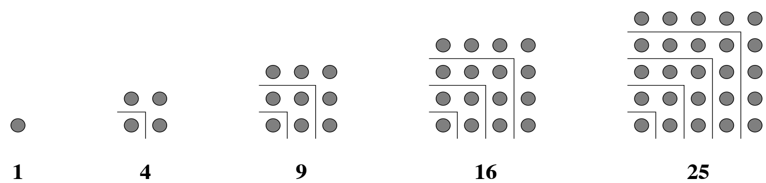


Figure 2.4: Square numbers. Applying the "rule" shows that consecutive square numbers are the partial sums of the odd numbers, 1, $1+3$, $1+3+5$, ...

squares can be seen to be the sums of the odds: $1 = 1$, $4 = 1+3$, $9 = 1+3+5$, and so forth. The oblong numbers can be arranged in figures in which one side exceeds the other by one unit. The oblong numbers are the sums of consecutive even numbers, as can be seen by applying the *gnomon* (Fig. 2.5).

Gnomon

Like the English word 'rule', the ancient Greek *gnōmōn* ($\gamma\nu\acute{\omega}\mu\omega\nu$) could refer either to an instrument that makes something known (such as a carpenter's square, a ruler, a quadrant, or the needle of a sundial), or more generally

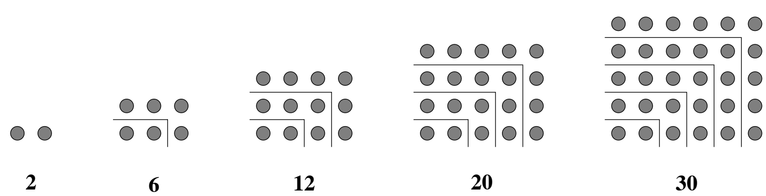


Figure 2.5: Oblong numbers. Application of the "rule" shows that consecutive oblong numbers are the partial sums of the even numbers, 2, $2+4$, $2+4+6$, ...

to a rule to be followed or to one who knows, judges or interprets. It is related to some of the words meaning *to know* (*gignōskō*), knowledge (*gnōsis*) and a rule or principle (*gnōmē*). Thus, concretely, the series of square and oblong numbers are generated by applying the shape of the carpenter's square, but more abstractly by applying an intelligible (*gnōrimon*) rule.[5]

Recursive Rules

The Pythagoreans were very impressed by the fact that these families of structures were generated by the *recursive* application of a single rule (Sinnige, *M&I*, p. 70). Recursive generation is still valued in science: sentences in the formal grammars used by linguists, logicians and computer scientists, definitions in mathematics, knowledge representation structures in AI and cognitive science — all of these make use of the recursive application of a finite number of rules to a finite number of terms.

"Figures"

In logic, AI and cognitive science, we often refer to a formal pattern as a *schema*, and it is no coincidence that this is the word (σχῆμα) the Pythagoreans used for the shape in which the pebbles were arranged. The corresponding Latin word, *figura*, is the origin of our word *figure*, and it is due to Pythagorean *figured numbers* that we still call numbers *figures* and refer to calculation as *figuring*. Both the Greek and Latin terms refer to the patterns or arrangements of things. Another term used to refer to the arrangement was Greek *eidos* (εἶδος), which comes from *to see*, and means appearance, aspect, form, figure, kind and so forth. Significantly, it was also used to refer to musical scales. A related word, *idea* (ἰδέα), is the origin of our word *idea* and is one of the terms Plato used to refer to his *forms* (Section 2.4). The latter is just the English derivative of the Latin *forma*, which has a similar meaning to the Greek *schema*. It is the basis of our notion of a *formal* system.[6] But also we see the roots of an assumption that *ideas* are formal structures, and hence that intelligence may be reduced to a formal system.[7]

Terms

[5]See Donnegan (*Lex.*), *LSJ*, Sinnige (*M&I*, pp. 70–75) and Peters (*GPT*). The Greeks got the *gnomon* — the needle of the sundial — from the Babylonians, who also arranged pebbles in right triangles for calculating their sides (Kirk, Raven & Schofield, *Presoc.*, pp. 83, 103, 335; see Neugebauer, *ESA*, Ch. II).

[6]See Burnet (*GPI*, pp. 49–53) and Taylor (*VS*, Ch. 5), as well as pp. 18 and 46.

[7]Pythagorean representation of numbers in *figures* may have been suggested to them by the constellations, and they probably knew that the Babylonians distinguished two aspects of a constellation: the number of stars in it and their arrangement (Maziarz & Greenwood, *GMP*, p. 13). The Babylonian view may have suggested to the Pythagoreans a distinction between the substance and the form of a thing, a characteristic feature of later Greek philosophy.

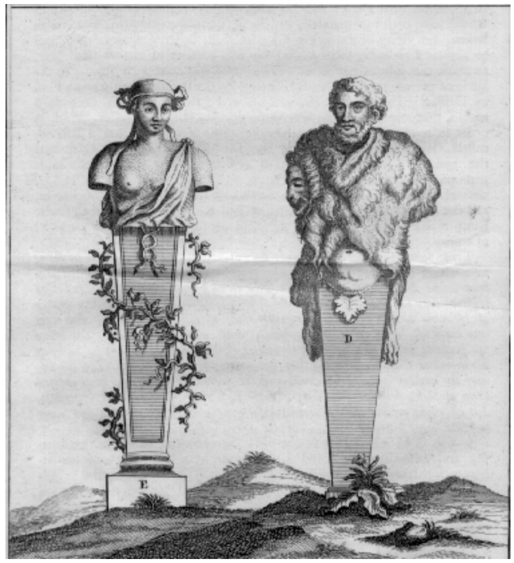


Figure 2.6: Termini or Boundary Markers.

The Pythagoreans called the pebbles in their figured numbers *boundary stones*, and called the spaces that they defined *fields*. However, the Greek word for these stones (*ὅρος*, *horos*) and its Latin equivalent, *terminus*, have a spread of meanings, including landmark, stone tablet, boundary, and, more abstractly, limit, standard, measure, aim, goal, rule and definition. These all connote definition or delimitation. This constellation of meanings is still with us in our *term*. We refer to a term in logic and mathematics, a technical term, to run to term, a school term, a prison term, terms of surrender or agreement, and speak of coming to terms with, and being on equal or good terms with. These all have connotations of mark, limit, measure or goal.

Sacredness of Boundaries

It is of course reasonable that in early agricultural societies the marking out of fields by boundary stones is fundamental to the structure of the society. They provided a definitive basis for resolving land disputes, and it is easy to imagine their becoming the principal metaphor for anything that is defining, delimiting, or conducive of order. As evidence of the importance of boundary markers, we find that in ancient Rome: "Offenses against the gods included murder, the slaying of a parent, incest, the selling of one's wife, the swearing of false oaths, and the moving of boundary stones, this last being a particular affront to the god Terminus" (Humez & Humez, *ABC*, p. 123). (See Fig. 2.6.) At one time, anyone pulling up such a stone could be killed with impunity and without the killer becoming defiled. The importance of Terminus is illustrated by the story that he was the only god that refused to give way to Jupiter when the latter came to reside on the Capitol. *Termini* (boundary stones, terms) were considered statues of the god and so were crowned with garlands and honored with sacrifices. Terminus was also celebrated in year-end festivals:[8]

> The simple neighbors meet and hold a feast, and sing thy praises, holy Terminus; thou dost set bounds to peoples and cities and vast kingdoms; without thee every field would be a root of wrangling. (Ovid, *Fasti* 2.657–660)

Boundaries and Geometry

Finally, recall also that *geometry* means *the measurement of land* and that we are told by Herodotus (*History* 2.109) that it had its origins in Egyptian surveying.[9] Thus both number theory and geometry have their origins in the dividing of continuous land. (See also Section 2.2.3 and Cohen & Drabkin, *SBGS*, p. 34.)

Significantly, the words *horos* and *terminus* were used to refer to the terms of a proposition or of a ratio. And there you have it. For the ancients *terms* were tokens that, by a recursive *rule*, could be arranged into *forms*, *figures* and *schemas*, and which thereby put knowledge in *rational* (or *logical*) form. This became the dominant root metaphor for knowledge for the next 2500 years.

Calculi and Calculation

It is well known that in ancient times small pebbles were used for calculation, for voting, and in various games. Indeed, it now seems that writing itself may have had its origin in the use of clay tokens for accounting (Schmandt-Besserat, ARS, EPW). In early Neolithic times, eleven thousand years ago, Mesopotamian merchants began to enclose tokens of various standard shapes in a clay envelope to indicate the contents of a shipment (i.e., a bill of lading). However, since the contents of the envelopes could not be

[8]See Guirand (*NLEM*, p. 214); Lemprière (*Cl. Dict.*, p. 615); Mollett (*Dict.*, pp. 315–316); Nettleship & Sandys (*Dict.*, pp. 620–621); *OCD*, p. 1045; Yonah & Shatzma (*Enc.*, p. 450).

[9]The impetus for the development of practical geometry may been the high population density in the Mediterranean region (Cornford, *FRP*, p. 142). Ancient tradition (Herodotus, Aristotle, Eudemos) held that geometry was brought to Greece by Thales (624–550 BCE). Although both the Egyptians and the Babylonians knew some practical geometry, its development as a logically structured science seems to have been initiated by Greeks, perhaps Thales himself (Ronan, *Science*, p. 68).

checked without breaking them open (which would be done only by the final recipient of the shipment), it was convenient to impress the shapes of the tokens on the outside of the envelope. Eventually, towards the end of the fourth millennium BCE, the enclosed tokens were omitted and their shapes were simply impressed on tablets, the predecessor of an ideographic writing system.

The *terms* of the Pythagoreans are of course another such use. The Latin word for such a token is *calculus*, and it is from the manipulation of calculi that we get our word *calculate*. We still use the word *calculus* for any game-like system in which *terms* are arranged in *schemas* and manipulated according to *formal* rules. The coincidence in terminology is not accidental, as we will see.

2.2.3 The Definite and the Indefinite

There is another issue in Pythagoreanism that we must discuss, for it sets the tone for much of Greek philosophy, and becomes a central issue in the foundations of mathematics and a motivation for symbolic knowledge representation in AI and cognitive science. It is related to the issue of boundary marking that we have already seen. The Greeks were uncomfortable when a continuum was not divided into discrete parcels by delimiting terms.

Finite and Infinite in Greek

The root, again, is the notion of a boundary, limit or end (πέρας, peras), but the more important term is *apeiros* (ἄπειρος), which is usually translated *infinite*.[10] More precisely it means *without internal or external limit*:

> Thus, in the context of a pre-Socratic philosopheme, and even still in Plato, *apeiron*, when translated into modern idiom, may have to be rendered by: infinite, illimited, unbounded, immense, vast, indefinite, undetermined; even by: undefinable, undifferentiated. (Bochner, Inf., p. 607)

Meaning of Finitus *in Latin*

It will be easier for us to understand the issue through Latin terms, since they are cognate to the relevant English words. The verb *finire* means to bound, limit, enclose within limits, restrain, determine, put an end to or conclude. The related noun *finis* means boundary, limit, border, term or territory. Finally, the perfect passive participle of *finire*, which is *finitus*,

[10] A compact discussion can be found in the entries in (Peters, *GPT*) for *apeiron* and *peras*.

Table 2.1: Pythagorean Table of the Ten Opposites

Limited	Unlimited
Odd	Even
One	Plurality
Right	Left
Male	Female
Resting	Moving
Straight	Curved
Light	Darkness
Good	Bad
Square	Oblong

means that which has been bounded, limited, restrained, ended, etc. This is of course cognate to our word *finite*, but is in fact broader in meaning. It means not just finite as opposed to infinite (endless), but also definite, determinate and limited. It includes the notion not only of a definite end, but also of clear and distinct boundaries and internal divisions. (Recall the importance of boundary markers, p. 25.)

It is hardly surprising that the Greeks considered the finite (in this broad sense) to be intelligible and good, whereas the infinite was chaotic and bad. "For evil belongs to the unlimited, as the Pythagoreans conjectured, and good to the limited" (Aristotle, *Nic. Ethics* 1106b29).

Table of the Ten Opposites

The Pythagorean preference for the definite is also expressed in their *Table of the Ten Opposites* (Table 2.1). These oppositions may be understood as follows.[11]

Limited vs. Unlimited

Limited vs. Unlimited This is of course the fundamental opposition upon which all the others are based. Indeed, all the others are mixtures, with some of the unlimited entering into the limited.

Odd vs. Even

Odd vs. Even The opposition is a bit obscure, but perhaps can be understood as follows (Aristotle, *Physica* 203a2). An even number can be divided or analyzed, but an odd number cannot. Therefore, so long as division yields

[11] Some of these explanations are ancient, but others are modern conjectures.

even numbers, analysis can continue; it is limited or ended by the reaching of an odd number. Therefore the odd numbers are the ultimate limits (or "atoms") of analysis. (We will see later that the notion of analysis stopping at "atoms" is fundamental to traditional epistemology.)

One vs. Plurality

One vs. Plurality Plurality contains an admixture of the void. For there to be discrete things there must be a principle of separation (void, gap = *chaos* in Greek).[12] In this sense the One is pure, unadulterated by *chaos*. The opposition of the One and the Many is a recurring theme in Western philosophy. The Pythagoreans held that

> the void distinguishes the natures of things, since it is the thing that separates and distinguishes the separate terms in a series. This happens in the first instance in the case of numbers; for the void distinguishes their nature.
> (Aristotle, *Phys.* 4.6.213b24–28)

(This observation is a deep insight into the topological distinction between the continuous and the discrete, as will be explained in volume 2.)

Resting vs. Moving

Resting vs. Moving The resting is *stable*, the moving unstable. The Greeks did not know how to make motion rational; the reduction of motion to ratios was not achieved until Galileo's time. The problem of change was central to all Greek philosophy, and much of it can be seen to be based on the assumption that change is inherently irrational (unintelligible). This is clearest in Plato (p. 45). Aristotle made the understanding of change central to his philosophy, but his theory was qualitative, i.e., he did not succeed in reducing change to ratios. We will see that the mathematical description of change, especially by the calculus, depended on a reconciliation of the rational and the irrational, the discrete and the continuous, the resting and the moving, the straight and the curved — all issues the Pythagoreans had identified.

Straight vs. Curved

Straight vs. Curved Straight lines have a constant (stable, dependable, determinate) direction; curved lines do not; since their direction is always changing, it is indeterminate. Also the length of a curved line is problematic,

[12] In Greek mythology, as in many others, the origin of the universe of discrete things comes with a separation of the undifferentiated continuum (Robinson, *IEGP*, Ch. 1).

as is the area under a curve. These notions were not clarified until the calculus was developed.

Light vs. Darkness

Light vs. Darkness The simplest explanation here is that in the light we see things clearly and distinctly, whereas in the dark everything is obscure and indeterminate.

Good vs. Bad

Good vs. Bad The only comment we make here is that it was a persistent theme of ancient Greek culture that the limited was good (cf. "Nothing in excess" on Apollo's temple at Delphi), and the absence of limits was bad (cf., the concept of *hubris*, or "overweaning pride").

Square vs. Oblong

Square vs. Oblong This is an unusual opposition, but it is important for the history of mathematics. The idea seems to be this (Aristotle, *Phys.* 3.4.203a10–15). The series of square numbers maintains a constant ratio of their sides, namely 1/1. On the other hand, for the oblong numbers this ratio is constantly changing: 1/2, 2/3, 3/4, 4/5, ... Of course we would say that this series approaches a limit, 1, but the concept of a limit was yet to be invented, and the Greeks had a hard time conceiving of the *limit* of an *unlimited* process. It was two thousand years before this problem was adequately resolved (if in fact it is yet). (See pp. 22 and 32.)

Other Oppositions

Other Oppositions The remaining oppositions (male vs. female and right vs. left) are difficult to understand in purely philosophical terms; they probably represent cultural biases of the Pythagoreans. For example, Wheelwright (*Pres.*, pp. 203–204) points out that constancy of direction was considered a masculine characteristic, and changeableness feminine. It must also be mentioned that in addition to their scientific endeavors, Pythagoreanism had a strong mystical component, and that they had many — to our mind — odd doctrines.[13] On the other hand, the association of the right side with the "propitious, healthy, strong, dexterous," and male, and conversely the left

[13]Dodds (*GI*, pp. 140–146) has argued persuasively that Pythagoras was a shaman, as were at least two more of the earliest Greek philosophers, Empedocles and Epimenides. It is perhaps no coincidence that we find these shaman-philosophers in late fifth-century and early sixth-century Greece, and that the Greeks' first contact with a culture based on shamanism came in the seventh century, when the Black Sea was opened to Greek trade and colonization. Against this, see Kirk, Raven & Schofield (*Presoc.*, p. 229).

with the "unfavorable, unsound, weak, … *sinister*" and female, has been a pervasive pattern in the Indo-European cultures (Mallory, *SIE*, p. 140).[14]

2.2.4 The Discovery of the Irrational

The foregoing discussion will perhaps make clear the devastating effect that the discovery of irrational numbers had on the Pythagorean brotherhood. It is possible that Pythagoras himself discovered the property that bears his name, and this led directly to the observation that the sides and hypotenuse of an isosceles right triangle are incommensurable. Let's try to understand this in terms of Pythagorean "figures." A ratio $m : n$ can be represented in a figure as a rectangle with sides m and n. What Pythagoras discovered is that there is *no* formula (arrangement of terms), no matter how big, that can represent this ratio exactly (Fig. 2.7). We can of course approximate it, but the exact ratio is forever beyond our grasp. Thus, although the hypotenuse surely has a length, it cannot be expressed by any (de)finite figure.

The implications of this discovery for the Pythagoreans was that their goal, which was to reduce all of nature to ratios, that is, to produce a rational account of nature, was doomed to failure. They had discovered a phenomenon of nature — in mathematics no less — which was, in their terms, by its nature irrational, and thus forever beyond the grasp of reason. This discovery destroyed the confidence expressed in "Everything is number."

Surds and the Absurd

Additional insight into the significance of this discovery on the Pythagorean outlook is provided by the etymology of the words *surd* and *absurd*. The word *surd*, in its mathematical sense of an irrational number, derives from the Latin *surdus* (deaf, inaudible, or insufferable to the ear), which is a translation of the Greek *alogos* (speechless, irrational). On the other hand, *absurd* origi-

[14] The Pythagorean Table of Opposites may be compared with the ancient (before 400 BCE) Chinese opposition of *yin* and *yang*: "Passive and active principles, respectively, of the universe, or the female, negative force and the male, positive force, always contrasting but complimentary. Yang and yin are expressed in heaven and earth, man and woman, father and son, shine and rain, hardness and softness, good and evil, white and black, upper and lower, great and small, odd number and even number, joy and sorrow, reward and punishment, agreement and opposition, life and death, advance and retreat, love and hate…" (Runes, *Dict.*, s.v. Yin and Yang). Other oppositions associated with yang and yin include light and dark, warm and cold, strong and weak, dynamic and passive, creative and receptive. For the most part the Chinese oppositions agree with the Pythagorean, although it is worth noting that in Taoist thought the yin (feminine) was considered preferable to the yang (masculine) (Schwartz, *WTAC*, p. 203; Laotse, *WoL*, Ch. 28).

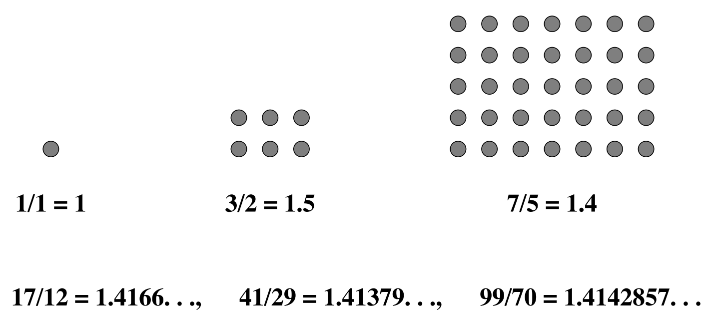


Figure 2.7: Figures Approximating Square Root of 2. The figures come closer and closer to expressing the ratio of the side to the hypotenuse ($\sqrt{2}$), but they never reach it. Hence, the relationship was considered irrational and the process infinite. See also the opposition of square and oblong numbers (p. 30). In this case the "rule" that generates the elements of the series is as follows: the height of the next figure is the sum of the width and height of the previous figure; the width of the next figure is the width plus twice the height of the previous. We have this procedure from Theon of Smyrna (fl. c. 115–140 CE), but it probably goes back to the early Pythagoreans (Heath, *Euclid*, Vol. 2, p. 119; Maziarz & Greenwood, *GMP*, pp. 121–122).

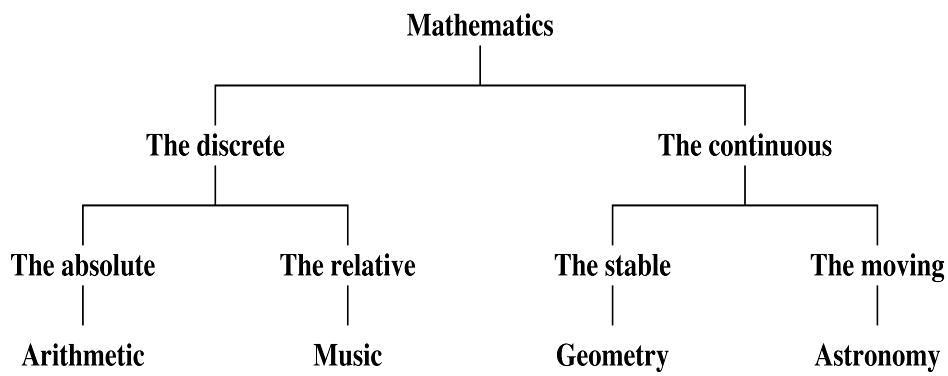


Figure 2.8: Pythagorean Divisions of Mathematics. From late antiquity through the middle ages, the four mathematical sciences were called the *Quadrivium*. Together with the *Trivium* — grammar, logic and rhetoric (which we might call syntax, semantics and pragmatics) — they made up the Seven Liberal Arts of the medieval schools.

nally meant inharmonious, jarring and out of tune (cf. Pythagorean musical theory, p. 22). It comes from *ab* (an intensive), and *surdus*. Thus, to the ancients it was nearly tautological that *surds were absurd*.[15]

2.2.5 Arithmetic vs. Geometry

The Rift Between Arithmetic & Geometry

The discovery of the irrational caused a major setback in mathematics at the end of the fifth century BCE (Maziarz & Greenwood, *GMP*, p. 5), and resulted in a split between arithmetic and geometry that was to last for two thousand years. On the one hand was the Pythagorean arithmetic calculus: the theory of natural numbers seemed like rationality in its truest sense. On the other hand, the demonstrations of the earlier geometers (perhaps Pythagoras himself) seemed convincing. Each of the two sciences, arithmetic and geometry, seemed to yield irrefutable laws, yet they remained unreconciled. As a result, mathematics split into two subdisciplines (Fig. 2.8),[16] and a major research problem in the philosophy of mathematics was born:

Future discussions will center around the 2 Pythagorean oppo-

[15] *American Heritage Dictionary* and *Oxford English Dictionary*: 'absurd' and 'surd'.

[16] H. W. Turnbull ("The Great Mathematicians," in: Newman, *WM*, p. 85) says we owe to the Pythagoreans this division of mathematics, as well as the word *mathematics* itself.

> sites of the indefinite (continuous) and the finite (discrete). But no synthesis of these two principles has yet been found to satisfy equally mathematicians and philosophers. (Maziarz & Greenwood, *GMP*, p. 65)

Arithmetization of Geometry

Most attempts at a unification of mathematics have tried to reduce geometry to arithmetic, since the calculus-like manipulation of terms in schemas according to formal rules has always seemed more rational. This *arithmetization of geometry* — the attempt to ground geometry in something like Pythagorean number theory — will be discussed in detail below (Chapter 5). Suffice it here to say that the arithmetization of geometry was not accomplished until the nineteenth century (by Dedekind and Weierstrass); the methods lead directly to the theory of computation.

2.3 Zeno: Paradoxes of the Continuous & Discrete

> Zeno's argument, in some form, have afforded grounds for almost all the theories of space and time and infinity which have been constructed from his day to our own.
>
> — Bertrand Russell

2.3.1 Importance of the Paradoxes

Split between mathematics and philosophy

After the discovery of the irrational in geometry, the Pythagoreans broke into two groups; one concentrated on mathematics, the other had more mystical interests.[17] Likewise, we shall, for a time, have to follow two parallel paths (they don't rejoin until the nineteenth century). On the one hand we have the history of mathematics trying to reconcile the discrete and the continuous; the only alternative would seem to be to abandon arithmetic or geometry. On the other hand, the second group of Pythagoreans clung to the idea that true knowledge is rational, but concluded that the forms are

[17] Dodds (*GI*, p. 67, n. 68) thinks this "split" is a modern fiction, imposing on the ancient Pythagoreans a modern dichotomy between science and mysticism. For the recency of this dichotomy, see Section 5.2.

not mathematical (where irrationality is inevitable), but more abstract. This is the path pursued by Socrates and Plato, which we will consider shortly. For now, however, we will follow the mathematical path a little further, and consider Zeno's paradoxes.

Zeno: c.490 – 430 BCE

Zeno's aim seems to have been to show that the continuous and the discrete are fundamentally irreconcilable, and in this he was quite successful. "The fact that it took 24 centuries to answer satisfactorily Zeno's arguments proves their fundamental importance in the history of mathematical philosophy" (Maziarz & Greenwood, *GMP*, p. 60). The formal apparatus of limits in modern mathematics makes it is easy to be glib about them, but, considered seriously, they still remain paradoxes. As Hamming (UEM) has said,

> Zeno's paradoxes are still, even after 2,000 years, too fresh in our minds to delude ourselves that we understand all that we wish we did about the relationship between the discrete number system and the continuous real line we want to model.

We'll see that the modern mathematical approach not without its own problems. The fundamental question of the continuous and discrete is: *In what sense in a continuum composed of discrete points?*[18]

2.3.2 Paradoxes of Plurality

As a defense of the thesis of his master, Parmenides, that "everything is one, altogether, changeless" (DK 28B8). Zeno proposed the following paradoxes to show the inconsistency of the idea that things are composed of units, as the Pythagoreans believed:

- The many have no size
- The many have infinite size
- The number of the many is finite and infinite

We'll consider each in turn.

The Many Have No Size

[18]An excellent discussion of Zeno's paradoxes can be found in Maziarz & Greenwood (*GMP*, Ch. 6). It is the basis for much of the presentation here. Additional information can be found in Kirk, Raven & Schofield (*Presoc.*, Ch. IX), Robinson (*IEGP*, Ch. 7), Burnet (*GPI*, §§ 63–66) and Sinnige (*M&I*, Ch. IV).

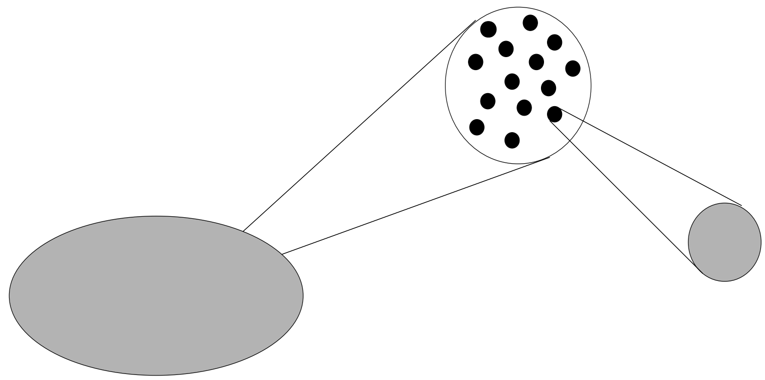


Figure 2.9: The Many Have No Size

> *The Many Have No Size:* "If it [the unit] existed, it would have to be one. But if it were one, it could have no body. If it had thickness, it would have parts, and then it would no longer be one." (Melissus of Samos, in Simplicius, *Phys.* 109.34; DK 30B9)[19]

The idea seems to be as follows (Fig. 2.9). Suppose that a thing is composed of units. Then these units must have no size. That is, they must be infinitely small (infinitesmal), since if they had any size, they would have parts (e.g. left and right sides). But such a unit doesn't exist at all, "for, having no size, it could not contribute anything to the size of that to which it was added. And thus the thing added would be nothing" (Simplicius, *Phys.* 139.5; DK 29B2). See also Robinson (*IEGP*, p. 129).

The Many Have Infinite Size

> *The Many Have Infinite Size:* "If they exist, each must have some size and thickness, and one part of it must project beyond the other. And the same argument applies to the projecting part; for this too will have size, and some part of it will project. Now to say this once is the same as saying it forever." (Simplicius, *Phys.* 140.34; DK 29B1)

The picture may be something like this (Fig. 2.10). If it has size, then it has

[19] The abbreviation 'DK' refers to the fragment's "Diels-Kranz number," its position in Diels & Kranz (*Frag.*). Freeman (*APSP*) provides a translation indexed by DK number, but Hussey (*Presoc.*, p. 156) claims it is unreliable and recommends instead Kirk, Raven & Schofield (*Presoc.*), Guthrie (*HGP*), Burnet (*GPI*) or Burnet (*EGP*).

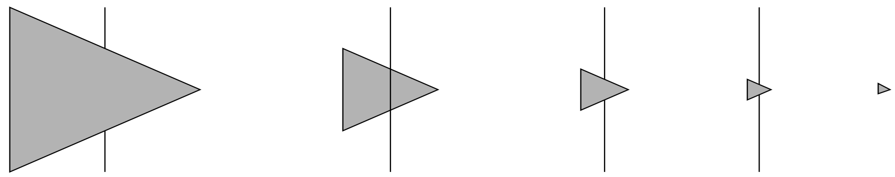


Figure 2.10: The Many Have Infinite Size

parts, but these parts also have size. And so we have an infinite number of parts, all with finite size.

The Number of the Many is Finite & Infinite

> *The Number of the Many is Finite & Infinite:* "If there is a many, there must be just so many — neither more nor less. But if there are just so many, they must be limited in number." That is, a (de)finite number. But, "If there is a many, there must be an infinite number of them. For between existing things there are always others, and between these others still." (Simplicius, *Phys.* 140.27; DK 29B3)

That is, an in(de)finite number. So again we reach a contradiction by assuming that there is a many, that is, that things are composed of discrete units.

2.3.3 Paradoxes of Motion

Zeno's paradoxes of motion can be organized as shown here:

	continuous	discrete
absolute motion	Dichotomy	Arrow
relative motion	Achilles	Stadium

They can be classified in terms of whether they're problems of the continuous or problems of the discrete. The Dichotomy and the Achilles are both problems of the continuous; they show the difficulties that arise when we assume space is infinitely divisible. We are left with an infinite number of pieces, all of finite size. If we think of them as discrete units then they seem to combine to an infinity (Fig. 2.11). The Arrow and the Stadium are both problems of the discrete; they show the difficulties that arise when we assume time is composed of discrete moments (Fig. 2.12). I'll discuss each paradox briefly.

The Dichotomy

Figure 2.11: Problems of the Continuous. The Dichotomy and the Achilles assume that space and time are infinitely divisible.

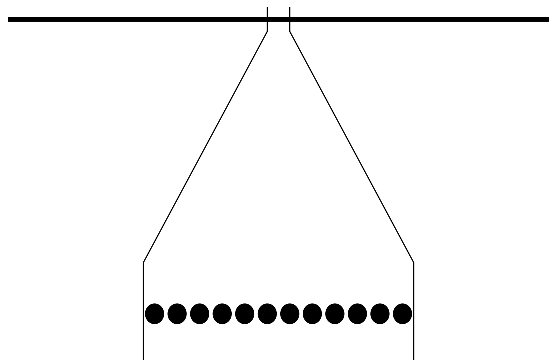


Figure 2.12: Problems of the Discrete. The Arrow and the Stadium assume that space and time are composed of indivisible units.

> *The Dichotomy:* "There is no motion, because what moves must arrive at the middle of its course before it reaches the end." (Aristotle, *Physics* 239b11)

That is, before we reach the point 1, we must pass through the point 1/2, and before we can do that we must pass through 1/4, and so on (Fig. 2.13). Hence we must pass through an infinity of points — each requiring finite time to reach — in finite time: $1/2 + 1/4 + 1/8 + \cdots$.

The Achilles

> *The Achilles:* "The slower in a race will never be overtaken by the quicker; because the pursuer must first reach the starting point of the pursued, so that the slower must always be some distance ahead." (Aristotle, *Physics* 239b14)

Suppose for simplicity that the slower is given a head start of 1 meter, and that the faster is twice as fast as the slower. By the time the faster has covered

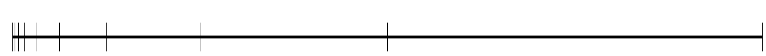


Figure 2.13: The Dichotomy

the 1 meter, the slower will have advanced another 1/2 meter. By the time the faster goes that 1/2 meter, the slower will have gone another 1/4, and so on. The slower will always be a little ahead of the faster. Of course Zeno knew as well as we do that the quicker will overtake the slower. The point of the paradox is to show a contradiction between this common experience and our theoretical reasoning about continuous motion and infinite divisibility.

The Arrow

> *The Arrow:* "The flying arrow is at rest"; because a thing is at rest when occupying its own space at a given time, as the arrow does at every instance of its alleged flight. (Maziarz & Greenwood, *GMP*, p. 59; cf. Aristotle, *Physics* 239b29, 5)

This is the problem of "instantaneous velocity." Suppose that at a given indivisible instant the arrow is moving. But if it moves it must be at different places at different times. But this has divided the instant (into a before and an after), which contradicts its indivisibility. Thus, in an indivisible instant the arrow cannot move; it's at rest. But if it's at rest at every instant of time, then it cannot move at all.

The Stadium

> *The Stadium:* This argument "supposes a number of objects all equal with each other in dimensions, forming two equal rows and arranged so that one row stretches from one end of a race course to the middle of it and the other from the middle to the other end. Then if you let the two rows, moving in opposite directions but at the same rate, pass each other, Zeno undertakes to show that half of the time they take in passing each other is equal to the whole." (Aristotle, *Physics* 239b33–240a2)

The argument seems to be this. We have three rows of objects of the same length. One row is stationary, the other two move in opposite directions. The initial configuration is shown in Fig. 2.14. Now consider the point in time when the two moving rows are both aligned with the stationary row (Fig. 2.15). When this occurs, the first unit in row B will have passed all the units in row C, but only half the units in row A. But rows A and C are the same length, so in a given period of time it has gone both the distance and half the distance. The contradiction arises from supposing that the units are indivisible. Then, in the time it takes B to pass one unit of A it will pass half of a unit of C, thus contradicting its indivisibility.

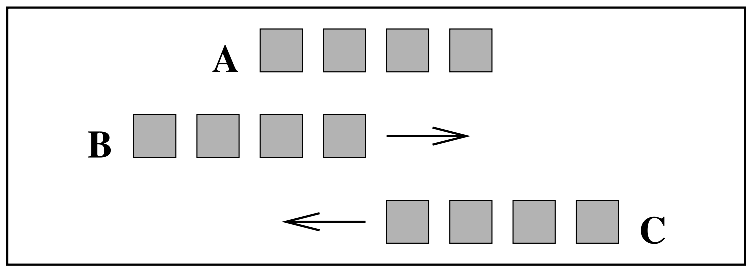


Figure 2.14: The Stadium: Initial Configuration

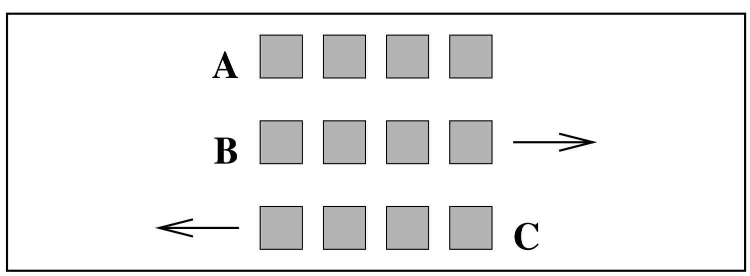


Figure 2.15: The Stadium: Rows Aligned

2.3.4 Summary

What has Zeno accomplished by these paradoxes? He has shown that if you assume that space and time are infinitely divisible continua, then you reach absurdities. On the other hand, if you assume that space and time are composed of discrete points or moments, then you also reach absurdities. Zeno's aim was to show that the notion of things having parts was incoherent and that, as Parmenides said, all is one. For our purposes though, the relevance of his paradoxes lies in the problems they reveal in the notion that a continuum is composed of discrete points. This problem is critical to the arithmetization of geometry.

2.4 Socrates and Plato: Definition & Categories

2.4.1 Background

Now we leave the mathematical path and consider Socrates' and Plato's development of Pythagorean mathematical and physical ideas into a theory of knowledge. They were so successful that they defined the theoretical framework for nearly all subsequent Western-philosophical debate about knowledge. In the epistemology of Socrates and Plato, word magic reaches a new level of sophistication.

Socrates, c. 470–399 BCE
Plato, c. 428–347 BCE

There is considerable doubt as to whether the ideas presented in Plato's dialogues are Socrates' own or Plato's. For our purposes, it doesn't much matter, since we will treat them as a unit. Plato, who was the most important student of Socrates, is one of the key figures in the history of philosophy. It has been truly said that Western philosophy is merely footnotes to Plato (Kaufmann, *PC*, Vol. I, p. 98). According to Burnet (*GPI*, Ch. IX), it is very likely that Socrates was a Pythagorean; Aristotle also thought his ideas were Pythagorean (*Met.* 987a–b). You may decide for yourself as we investigate his views.[20]

[20] Needless to say, there is an enormous literature on Socrates. I. F. Stone's 1988 book provides a nice overview of his philosophical ideas and how they led to his execution. This is perhaps not a majority opinion among scholars (Griswold, SGP), but I find it convincing. More traditional views are presented in Brickhouse & Smith (*SoT*). Burnet (*GPI*, Chs. 8–10) has an interesting account of the historical Socrates, which emphasizes the Pythagorean connections, although, again, this position is considered extreme by many

2.4.2 Method of Definition

Facts vs. Words

We have seen the importance to the Pythagoreans of *logoi*: ratios, terms, words, and rational accounts. Therefore, Socrates' probable Pythagoreanism will explain the importance he attached to words. In fact, a shift of emphasis from facts to words was the essence of his contribution to philosophy:

> We know from Plato that the new method of Sokrates consisted precisely in the consideration of things from the point of view of propositions (*λόγοι*) rather than from that of facts (*ἔργα*) ... (Burnet, *GPI*, p. 146)

An important example of this is his Method of Definition, which is based on the belief that we do not understand something unless we can define it, and that therefore definition should be the principal activity of philosophers. The idea is essentially Pythagorean: to define means to make something definite, and to make it definite is to bound it and set it off from other things. Recall the Pythagoreans' concern with the (de)finite and the in(de)finite. For the ancient Greeks, to be intelligible was to be definite (and hence defined).[21] You can see why definitions would be so important to a Pythagorean like Socrates. Thus, many of the dialogues have as their goal the definition of such terms as *excellence*, *courage*, and *piety*:

Illustrative Quotations From the Dialogues

> ... what is that common quality, which is the same in all these cases, and which is called courage? (*Laches* 191e)
>
> Well then, show me what, precisely, this ideal is, so that, with my eye on it, and using it as a standard, I can say that any action done by you or anybody else is holy if it resembles this ideal, or, if it does not, can deny that it is holy. (*Euthyphro* 6e)
>
> And so of the excellences, however many and different they may be, they all have a common nature which makes them excellences. (*Meno* 72)

The emphasis on definition continues in philosophy to the present day. Most knowledge representation schemes in AI and cognitive science are likewise based on formal structures that represent a concept in terms of its defining properties ("that common quality" or "common nature").

---

scholars.

[21] See also Section 2.2.3 and p. 47.

2.4.3 Knowledge vs. Right Opinion

Socrates' entire theory of knowledge is centered on words, for he claimed that we truly know something only when we can give a verbal account of it. The Pythagorean orientation is apparent: something is rational or logical only when it can be expressed in terms of ratios and *logoi* (words, propositions, verbal accounts). As he says in the *Laches* (190c), "that which we know we must surely be able to tell." (See also *Meno* 96d–100a.) Of course Socrates recognized that many people are skillful in their endeavors, and yet unable to explain what they're doing in theoretical terms. Yet he denigrated this atheoretical, practical knowledge, and called it (merely) "right opinion." Such people, he said, knew what to do, but not why they should do it. He contrasted this with theoretical knowledge, which for him was the only true knowledge:

> it is not an art[22] [*techne*] but a practice [*empeiria*], because it can produce no principle in virtue of which it offers what it does, nor explain the nature thereof, and consequently is unable to point to the cause of each thing it offers. And I refuse the name of art to anything irrational.[23] (*Gorgias* 465a)

An art, as opposed to a practice, "has investigated the nature of the subject it treats and the cause of its actions and can give a rational account of each of them" (*Gorgias* 501a). For a concrete example, consider tuning a lyre. A musician can do it, but doesn't know why his technique works. He doesn't have true knowledge. Pythagoras, on the other hand, can give a rational account (in all senses of rational).

2.4.4 The Platonic Forms

The Socratic/Platonic theory of "forms" has been one of the most influential epistemological theories in Western philosophy.[24] It is most comprehensible when seen as an outgrowth of Pythagorean mathematics.

The Objects of Mathematics

[22]'Art' (*technē*) must be taken here to mean a systematic or methodical craft, or even an applied science; on the other hand, a 'practice' (*empeiria*) is based on experience or practice (*LSJ*, s.vv. τέχνη, ἐμπειρία; Peters, *GPT*, s.v. *'technē'*).

[23]N.B. our discussion of *irrational*, p. 21.

[24]The theory of forms is discussed in many of the Platonic dialogues. The following are a few key sources: Approximations to an ideal: *Phaedo* 74a–75d; How being and becoming

In ancient times — as now — it was held that the truths of mathematics are the most certain truths of all. Two plus two is exactly four; it's not possible that refined measurements will show it's 4.00001, and it's not possible that new discoveries will require this law to be rejected. Other examples of mathematical truths are the Pythagorean theorem, and the theorem that the angles of a triangle add to two right angles. But even if we grant the certainty of these truths — that they are *necessary* truths — we may still question what they are about. They're about numbers or triangles you say? But what is a triangle? Surely not the triangle we draw, which can never have perfectly straight edges, or be made of edges with no width. But these are the only triangles that exist, in the sense that physical objects exist. We may say that mathematical truths are about "idealized" triangles, which are products of thought. But it's clear that the truths of mathematics are objective; all rational investigators will find the angles of a triangle to be two right angles. Hence the triangles of mathematics must have an existence that is not physical, and yet is independent of individual mathematicians. Thus it seems that the only explanation for the objectivity of mathematics is that there is a "realm" where there exist the true, perfect, ideal lines, points, triangles, and other objects of mathematics. The mathematician explores this realm by a process of pure reason.[25]

Role of Drawings in Geometry

But if mathematicians are exploring the realm of ideal mathematical objects by pure reason, then why do they draw the figures and constructions that are so prominent in mathematical proofs? Plato's answer (*Republic* 510d–e) was that these are merely aids to the intuition. True intelligence passes beyond the need for these crutches and can proceed by reason alone.

Approximation to an Ideal

The example of mathematics is easily extended. The physical triangles in the everyday world of sensation are approximations to the ideal triangles that the mathematician studies. Similarly, when we say that two objects are equal, we recognize that this equality is an approximation to mathematical (perfect)

---

are known: 78d–80b; Philosophy frees us from senses: 82d–83e; How the forms are related to sensible objects: 99d–102d; A summary: *Parmenides* 128e–130a; Critique of the theory: *Parmenides* 130a–135c; Investigation of nature of knowledge: *Theaetetus*. Plato explains the realms of being and becoming in three famous "myths": *Republic* 506d–511e, 514a–519b.

[25] We are running roughshod over many important issues in the philosophy of mathematics, only a few of which will be treated later. The nature of mathematical objects and mathematical truth are still controversial topics. A good reference is (Benacerraf & Putnam, *PM*).

Equality. Furthermore, Plato claimed that the triangles and equalities of sense can be understood only by reference to the corresponding ideals. Now, since the Pythagoreans already believed that everything is number (p. 31), it's not such a big step to see the transient and imperfect virtues of individual people or things as approximations to an eternal, idealized and perfect Virtue that exists in the same realm as the mathematicians' triangles. Also, it's not such a big leap to say that these individual virtues can be understood only by reference to ideal Virtue. Philosophers, like mathematicians, are after eternal certainties, and so they investigate the ideals by pure reason. The objects of sense may prod the intuition, but ultimately they mislead.

Rationalism

One effect of this view has been the prevalence in early Western philosophy of *rationalism,* the view that pure reason is a much surer way to the truth than empirical investigation. 'Rationalism' is not a synonym for 'reasonableness'; rather it is a technical term referring to

> (a) the belief that it is possible to obtain by reason alone a knowledge of what exists; (b) the view that knowledge forms a single system, which (c) is deductive in character; and (d) the belief that everything is explicable, that is, that everything can in principle be brought under the single system. (Flew, *DP*, s.v. 'rationalism')

In this sense, rationalism is not the same as the practice of being rational, in the sense of being reasonable. Indeed, a significant question is whether rationalism *is* reasonable. Rejection of rationalism was a major feature of the scientific revolution in the sixteenth century. In a broader sense, Plato's views lead to *intellectualism,* the view that theoretical knowledge is the only true knowledge, and that so-called practical knowledge is "mere opinion" (Section 2.4.3). Intellectualism was not questioned by the scientific revolution, and it is a major background assumption of traditional AI and cognitive science, which tend to focus on intellectual and verbal skills to the exclusion of manual and other nonverbal skills.

Being and Becoming

The Pythagorean influence is very apparent in the Platonic distinction between Being and Becoming. Recall that the Pythagoreans consigned motion and change to the Indefinite (p. 29). Change was intelligible only when it could be reduced to ratios. Zeno's paradoxes of motion only reinforced this assessment (Section 2.3.3). Yet in the everyday world of sense, things are always changing; everything is in a state of becoming. Thus, the world of sense is in a fundamental way unintelligible, and can be understood only to the extent that it approximates the eternal (changeless) ideals in the world of Being. We can never have scientific knowledge about becoming; knowledge is always of being (p. 43).

Being, Becoming and Approximation

The notion of approximations to an ideal is connected with the distinction between being and becoming. The approximations are "striving" or "tending" to *become* the goal, but they will never *be* it. This is illustrated in Zeno's paradoxes of motion (Section 2.3.3). As Burnet (*GPI*, p. 156) says, "The problem of an indefinite approximation which never reaches its goal was that of the age." But a theory of limits did not come for two millennia.

The Term Form

The foregoing ideas are brought together in the theory of forms, but before I discuss it it's necessary to discuss terminology. The Greek words here translated *form* are εἶδος (*eidos*) and ἰδέα (*idea*; the source of English *ideal*).[26] These words are often translated *idea* (and thus one hears of Plato's Theory of Ideas), but that is a poor translation, since Plato's "ideas" are definitely not in the head. These words originally meant the form of a thing, its shape, or figure. It is significant that these words were also used to refer to the Pythagorean figures. This is evidence for the view that the theory of forms is a development of Pythagoreanism. Later these words came to mean a characteristic property or category. Notice the continuing assumption that categories are *formal*. (Recall also the discussion on p. 24.)

The Theory of Forms

In Plato's theory of forms two realms are postulated: the familiar realm of sensible objects and the realm of the forms. The realm of sense is characterized by flux and approximation. It is intelligible only to the extent that the sensible objects approximate the ideal forms. The forms themselves are changeless, ideal and perfect. Perception is a faulty source of knowledge; it informs us of the world of sense, which is unintelligible, and can at best hint at the forms. Knowing the forms requires pure reason. Reason is capable of comprehending the forms because the categories of thought are in fact the forms. The words we use for these categories (triangle, equality, virtue, etc.) are the names of the forms. True knowledge is thus knowledge of the forms and their logical relations.

Epistemological Implications

Since the forms correspond to what are commonly called categories and concepts, we can draw from the theory of forms the following conclusions about categories and conceptual knowledge:

[26] See Peters (*GPT*, pp. 46–47), Taylor (*VS*, Ch. 5), *LSJ* (s.v. εἶδος, ἰδέα), Donnegan (*Lex.*, s.v. εἶδος, ἰδέα), and Burnet (*GPI*, pp. 49–53).

Categories are real because they exist in the world of forms. Therefore there is nothing arbitrary or subjective about them.

Categories are static since there is no change (becoming) in the world of being. Therefore, categories do not evolve.

Categories are a priori because they exist independent of experience; they are not derived from experience.

Categories are context-independent because they fit into an eternal logical structure.

Categories are discrete because they correspond to *terms* (words), and terms are discrete.

Words have definite meanings because each word names a form, which is definite.

Categories have an objective logical structure. That is, the logical relations between categories are like those between the mathematical objects.

Terms have objectively correct definitions because the definitions are determined by the logical structures of the forms that the terms name.

All knowledge is formal knowledge because the only true knowledge is knowledge of the forms and their logical structures.

What we truly know we can say because words correspond to forms, and the logical structure of language — properly used — reflects the logical structure of the forms.

These assertions have been assumed — almost without question — throughout most of Western intellectual history, but especially in epistemology, cognitive science and artificial intelligence. Thus it is especially significant that they are rejected by *connectionism*, the new theory of knowledge which is the subject of the second half of this book.

2.4.5 Summary: Socrates and Plato

We have presented — very briefly — what is probably the most influential theory of knowledge and concepts in Western philosophy. In effect it provides a justification for the Pythagorean program. If the only true knowledge is knowledge of the forms, and if the forms are real discrete objects fitting into a logical structure, then such knowledge can be expressed verbally, as terms arranged in formal structures. Thus the truly real world, the world of Being, has a rational structure, even if the sensible world, the world of Becoming, which is only a distorted shadow of true reality, is ultimately irrational and unintelligible. The principal task of philosophy and science thus becomes the charting of the formal structure of the world of forms.

2.5 Aristotle: Formal Logic

2.5.1 Background

Aristotle: 384–322 BCE

Aristotle is one of the most influential thinkers in Western philosophy. If he stands behind Plato it is only because he was a student of Plato, and thus is the principal "footnoter" of his teacher (p. 41).[27]

Aristotle's scholarship had enormous breadth: he wrote on nearly every subject from logic, physics and biology, to love, music and table manners. He was nothing if not prolific: one ancient catalog[28] lists 150 books (about 50 modern volumes) comprising 445 250 lines! And this catalog is known to be incomplete! (Barnes, *Aris.*, p. 3) Unfortunately, only about a fifth of Aristotle's writings have survived the accidents of time and the hands of the book burners. It is no wonder that throughout most of history Aristotle has been known simply as "The Philosopher." Here we will be concerned only with Aristotle's logical works; these are the ones that have been most influential in the traditional theory of knowledge.

2.5.2 Structure of Theoretical Knowledge

Knowledge vs. Opinion

Recall Socrates' distinction between knowledge and right opinion (p. 43). Knowledge is preferable because it's more reliable. That is, if we just have

[27]Two readable summaries of Aristotle's philosophy are Randall (*Aris.*) and Barnes (*Aris.*). There are many books of selections from Aristotle's works.

[28]Diogenes Laertius, 5.22–27

right opinion, then we only know that what we are doing has worked in the past; we cannot be certain that it will work the next time we try it. On the other hand, if we have knowledge, then we can give a rational account of what we do. Therefore, since we know the necessary connections between things we do not have to fear being wrong.

Aristotle accepts this same basic definition, since he too expects true knowledge to be universal.[29] Also, like Plato, he sees that the only way to achieve this universality is to give knowledge a strict deductive (logical) structure grounded in indubitable premisses.[30] This leads to two subgoals in Aristotle's investigation of the structure of knowledge: one is to set down the rules for deductive argument; the other is to determine how we can know the primary truths, since these cannot be established deductively. We will discuss the results of Aristotle's investigations in each of these areas.

2.5.3 Primary Truths

Nature of Primary Truths

The primary truths must be more than mere assumptions, since in that case the conclusions drawn from them would be no better than assumptions. Further, the primary truths cannot be merely hypotheses, since then the conclusions would be no surer than hypothetical. For scientific knowledge to be absolutely certain, the primary truths themselves are required to be absolutely certain. But since the primary truths are the starting point of deductions they cannot themselves be established deductively. Therefore the primary truths must be *self-evident*, in the literal sense of providing their own evidence. That is, the primary truths are self-justifying.

Nature of Definitions

In most cases Aristotle takes the primary truths to be definitions or parts of definitions. But again we must be careful, since Aristotle understands definition differently from the way we do now. In the deductive sciences we usually take a definition to be a *prescription* for the use of a word. That is, a definition is way of introducing a word as an abbreviation for a longer

[29]Aristotle distinguishes three kinds of scientific knowledge (*Metaphysics* 6.1.1025b25): theoretical knowledge, productive knowledge and practical knowledge. At the present time we are concerned only with theoretical knowledge, and when we use the term *knowledge* this is what it will mean.

[30]Aristotle, like Plato, was an epistemological "realist," which means that he took the forms to exist independently of us and not to be creations of our minds. He differed from Plato in putting the forms in the objects of sensation rather in a separate ideal realm. However, these ontological distinctions are not relevant to our concerns here.

sequence of words. Such definitions are not truths, but conventions, and therefore would lead to no more certainty than arbitrary assumptions.

The modern notion of definition will not serve Aristotle's needs. For him a definition is a factual statement that says what it is for a thing to be what it is. An example will make this notion clearer.

Example Definition: Triangle

We can begin with a good example of a definition: 'a triangle is a three-sided figure'.[31] The purpose of this definition is *not* to introduce 'triangle' as an abbreviation for 'three-sided figure', nor is it even to explain the way the word 'triangle' is used in English. Rather, its purpose is to state what it is for something to be a triangle. As it's usually put, the definition states the *essence* of triangles: the properties that anything must have in order to be a triangle. Something that's not a figure, or that's not three-sided, is surely not a triangle. Conversely, any figure with three sides is surely a triangle.

Essence *in Aristotle*

Traditionally, definition in terms of essences is considered the hallmark of Aristotle's theory of definition, and much medieval (and even modern) philosophy was concerned with the nature of essences. Yet it's remarkable that there is not a Greek word corresponding to the translation *essence*. The phrase most commonly translated *essentially* is *καθ' αὑτό* (*kath' hauto*), which means *per se*, or *in itself*. So where we often read "what is Man essentially?" or "what is Man in essence?", we should read "what is Man in itself?" Similarly, there is no single word corresponding to *essence*. The phrase most commonly translated this way is *τὸ τί ἐστι* (*to ti esti*), which means *the 'What is it?'* (a question turned into a noun). Another such phrase is *τὸ τί ἦν εἶναι* (*to ti ēn einai*), which means something like *the 'What is it to be what it is?'*. These translations are more awkward, but more accurate (Randall, *Aris.*, p. 47, n. 13). We will avoid *essence* and derivative terms.

Influence of Language on History of Ideas

The problem of *essence* is a good illustration of the role of language in the history of ideas. The Latin *essentia* was coined, perhaps by Cicero, to translate the Greek *ousia* (one's own, property, being); in the Medieval period it came to mean *essence* in the sense under consideration here.[32] Over the two millenia since its invention, much ink has been spilled about essences — what they are, where they are, and so forth.

[31] The definition we use for this example, 'a three-sided figure', admits triangles bounded by curved lines, and these are traditionally called triangles. However, for the purpose of the example we restrict our attention to *rectilinear* figures and triangles, that is, those bounded by straight lines.

[32] At least as early as Thomas Aquinas (1225–1274), e.g., *Sum. Theol.* 1, q.3, a.3 concl. and q.29, a.2 ad 3.

But observe: by creating a word, Cicero (or whoever) created a philosophical problem. Once the word *essentia* had been invented and used in meaningful contexts, such as translations and paraphrases of Aristotle, it was necessary to find something that it named. The implicit presumption is that if it can be used meaningfully, then it must mean *something*, in other words, there must be *some things* (essences) to which the word refers.

The existence of a word such as *essence* can also bias the way we go about our investigations and can set bounds on acceptable answers. If we begin our inquiry by seeking "the essence of life," we will likely find the soul or an *élan vital* or some such. On the other hand, if we begin by asking, "What is it to be alive?" then we are more likely to come up with a description of a process, or at least an operational test for life. Thus we must beware of the "bewitchery of words."

Back to Aristotelian definitions. Since they state the most fundamental properties of things, their discovery may require significant analysis and scientific investigation. Once found, however, they are self-evident in the way illustrated above. Who could rationally deny that a triangle is a three-sided figure?

Primary Truths in Definitions

A proper Aristotelian definition contains one or more primary truths. For example, in the definition of triangle we may see two primary truths: that a triangle is a figure, and that a triangle has three sides. From these and other primary truths many derivative truths follow in turn by deduction. In summary, Aristotelian definitions are *self-evident matters of fact*, not prescriptions.

Independence of the Sciences

There is one more aspect of Aristotle's approach that we must address before leaving the topic of primary truths. This is that Aristotle permits the various sciences to have their own primary truths; he does not seek to derive all truths from one first principle, as Plato did. As we'll see, Euclid, following Aristotle, deftly avoids the chasm between arithmetic and geometry — the discrete and the continuous — by basing each science on its own primary truths (p. 55).

Difficulty of Discovering Definitions

Unfortunately, it's much more difficult to apply Aristotle's idea of definition outside of mathematics. What is the definition of cow, or person? To define person we must find those properties without which a thing would not be human.

2.5.4 Formal Logic

Analysis of Propositions into Terms

Although many earlier philosophers had studied the forms of arguments (especially the Sophists), we owe to Aristotle the founding of logic as a science. He was the first to analyze propositions into *terms* and to show how deductive processes rearrange these terms (recall p. 24). For example, consider the well-known syllogism:[33]

	All men are mortal;
	Socrates is a man;
therefore,	Socrates is mortal.

The validity of this argument does not depend on the particular terms 'Socrates', 'man' and 'mortal' that appear in it; indeed, they are like game tokens (calculi, p. 26). All that's important to the validity of the argument is its *form* (hence, *formal* logic).

Formal Rules

The general form of this argument can be expressed in a formal rule, or *schema*, such as this:

	All M is P
	S is M
therefore,	S is P

Indeed, Aristotle was the first to use variables (such as S, M and P here) to express rules formally; it is a major contribution and a model for rule-based systems in AI and cognitive science.

Figures of the Syllogism

Aristotle considered all the possible arrangements of the terms in syllogisms and classified them into three *figures* (*schemata*, p. 24). The preceding example is in the first figure; here is a valid syllogism in the second (Joseph, *IL*, p. 258):

	No insects have eight legs;
	Spiders have eight legs;
therefore,	Spiders are not insects

In general:

[33] I retain the conventional translation 'man' for ἄνθρωπος (*anthrōpos*), which, though masculine in gender, was generally used for people of both sexes.

$$\begin{array}{rl} & \text{No } P \text{ is } M \\ & S \text{ is } M \\ \hline \text{therefore,} & \text{No } S \text{ is } P \end{array}$$

The three figures enumerate the possible arrangements of the three terms that occur in the syllogism: S the subject of the conclusion, P the predicate of the conclusion, and M the *middle term*, which appears in both premisses but not in the conclusion. Writing the terms of the propositions in the order subject-predicate, we have the three figures:

$$\begin{array}{ccc} MP & PM & MP \\ SM & SM & MS \\ \hline SP & SP & SP \end{array}$$

Note that this exhausts all possible arrangements, if the order of the premisses is not considered.

Knowledge as a Formal Structure

Aristotle's formal logic can be considered a continuation of the Pythagorean program. The earliest Pythagoreans thought that things were literally composed of numbers, that is, units (terms) arranged in various forms or figures. Later Pythagoreans believed a more abstract version of this theory: that every thing had a number through which it could be understood. Aristotle moves to a higher level of abstraction, since for him it's not things that are formal arrangements of terms, but knowledge itself. What has not changed is the identification of the intelligible with formal structures.

2.5.5 Epistemological Implications

We now turn to some of the epistemological implications of Aristotle's view. Since for Aristotle definitions are matters of fact, there is one correct definition for each term, that is, the definition is a formula (*logos*) saying "what it is to be what it is." Like Socrates and Plato (Section 2.4.2), Aristotle believed that the meaning of a term can be expressed exactly in a finite formula.

Similarly, as we've seen, Aristotle was able to express his deductive rules formally — as mechanical symbol manipulation processes. Therefore, in Aristotle's ideal of a completed science, all the knowledge is expressed as formal (structural) relationships between symbol structures (schemata, formulas).

We summarize the epistemological implications of Aristotle's theory:

- Definitions are objective matters of fact, which can be expressed in finite formulas.
- Deduction can be described by the formal manipulation of terms arranged in specified schemata.
- A completed science takes the form of propositions connected formally to definitions.

These assertions have become incorporated into our unconscious assumptions about "true knowledge," and they provide the ultimate source of the formal, deductive knowledge representation and inference schemes commonly employed in cognitive science and artificial intelligence. However, in volume 2 we will see that they are assumptions that need to be questioned, and in fact rejected.

2.6 Euclid: Axiomatization of Continuous & Discrete

> Euclid alone has looked on Beauty bare.
>
> — Edna St. Vincent Millay (*The Harp Weaver*, 4, sonnet 22)

2.6.1 Background

Eudoxus: 408–355 BCE

Euclid: c. 300 BCE

We return now to the mathematical part of our story, and consider an important investigation of the continuous and discrete in mathematics. Eudoxus, a student of Plato, was probably the greatest Greek mathematician before Archimedes.[34] It is likely that he originated both the theory of magnitudes and the method of exhaustion, which we find in Euclid's *Elements*. Yet not one of his works survives (Bochner, *RMRS*, p. 325). On the other hand, by all accounts (ancient and modern) Euclid was a rather mediocre mathematician. Nevertheless, the 13 books of his *Elements* have survived intact,

[34] A general source for the material in this section is Maziarz & Greenwood (*GMP*, Part 4). The definitive translation of Euclid's *Elements* is Heath (*Euclid*).

and have been a required subject in school from his time until well into the twentieth century. Its apparently perfect reduction of a body of knowledge to a deductive structure has an austere beauty, as Millay and many others have recognized.

2.6.2 Axiomatic Structure

Euclid & Aristotelian Science

Euclid's *Elements* is an application to mathematics of Aristotle's idea of a science as defined in his two major logical works, the *Prior* and *Posterior Analytics* (Maziarz & Greenwood, *GMP*, p. 242–243). It begins with definitions in terms of necessary and sufficient attributes that are taken to be prior to the term defined. It bases its deductions on axioms ("common notions"), which are taken to be self-evident truths, and postulates, which are taken as the starting points of the particular science (*Post. An.* 74b5–77a30). The organization of the whole makes its deductive structure explicit, since no proposition may be admitted unless it is deducible from the first principles.[35] The *Elements* was thus the first concrete demonstration of how a body of knowledge could proceed by formal operations from explicitly given hypotheses. It remained the exemplar of formal reasoning until some of its defects were discovered in the nineteenth century.

Knowledge Elements

The Platonic/Aristotelian view of knowledge as a formal structure of discrete propositions is further evident in the use of the term *elements*. Pre-Euclidean mathematicians had already organized theorems by showing that many of them followed from a few general principles, which they called *elements*, by analogy with the alphabet's relation to language (Maziarz & Greenwood, *GMP*, p. 240). Compare Plato's notion of the unanalyzable *elements* of which the "syllables" of knowledge are composed (*Theaetetus* 201d–206b). In both cases there is a presumption that knowledge is a complex of discrete, indivisible elements. This view is characteristic of the traditional view of knowledge, as we will see (Sections 4.3.2).

[35] Aristotle's use of rational necessity should be contrasted with Plato's. Plato used rational analysis as a means of regression from the familiar forms back to the most basic form. Aristotle takes certain principles as given, and the then by rational synthesis shows how various conclusions follow from them by rational necessity (Maziarz & Greenwood, *GMP*, p. 242–243).

2.6.3 Theory of Magnitudes

The Pythagoreans and Zeno had demonstrated the difficulty of having a single theory that encompasses both discrete and continuous quantities. Therefore, Euclid axiomatized each of them separately. In Book 7 he develops the theory of discrete quantities — what we call number theory. However, in Book 5 he develops an axiomatic theory of continuous quantities, or *magnitudes*. This is based on relations of proportion, that is, on *ratios*. Using this theory he is able to prove the very important *principle of continuity*, which is the basis for the *method of exhaustion* — a way of finding the limits of sequences (p. 57). This principle shows that certain infinite series must eventually get smaller than any number we can pick. We consider briefly Euclid's theory.

Magnitudes vs. Numbers

Just as numbers (i.e. integers) are idealizations of discrete objects, such as pebbles or tokens, taken as members of ensembles, so magnitudes are idealizations of continuous quantities, such as lengths and areas. Both idealizations are based on intuitions about the familiar world. For example, we see we are surrounded by discrete objects. We also see continuous change, such as continuous motion, growth, and the flow of time.

Numbers More Basic than Magnitudes?

Although we have basic intuitions of both the continuous and the discrete, our Pythagorean view of knowledge has caused us to view numbers as more basic than magnitudes — hence the goal of *arithmetizing geometry* (Chapter 5). We will see in volume 2 that we can as easily *geometrize arithmetic*, that is, reduce the discrete to the continuous.

Formality of Theory of Magnitudes

Euclid's theory of magnitudes is not expressed with nearly as much rigor as would be demanded now. In contrast to the axiomatization of geometry in Book 1, where point, lines, and the like are defined, the basic concept *magnitude* is not defined at all. The definitions we find in Book 5 have to do with multiples, ratios, proportions, and so forth. Further, there are no postulates for magnitudes. Rather, the proofs are based on the axioms (Common Notions) from Book 1 together with informal intuitions about magnitudes.[36]

Principle of Continuity

The Principle of Continuity, which is the basis for Euclid's method of handling limits, is of fundamental importance for the eventual arithmetization

[36] The Common Notions are: (1) Things which are equal to the same thing are also equal to one another. (2) If equals be added to equals, the wholes are equal. (3) If equals be subtracted from equals, the remainders are equal. (4) Things which coincide with one another are equal to one another. (5) The whole is greater than the part. (Euclid, Bk. 1)

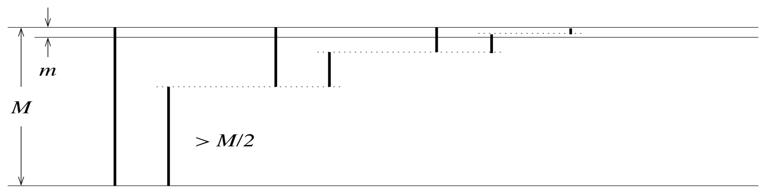


Figure 2.16: Principle of Continuity. M and m are two unequal magnitudes, $M > m$. Subtract from M a magnitude $> M/2$ and consider the remainder. Subtract at least a half of the remainder, and continue. Eventually a magnitude smaller than m will remain.

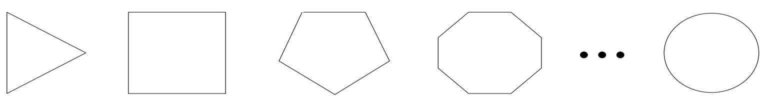


Figure 2.17: The Circle as an Infinite-Sided Polygon

of geometry:

> Two unequal magnitudes being set out, if from the greater there be subtracted a magnitude greater than its half, and from that which is left a magnitude greater than its half, and if this process be repeated continually, there will be left some magnitude which will be less than the lesser magnitude set out . . . And the theorem can be similarly proved even if the parts subtracted be halves. (Euclid, Bk. 10, Prop. 1)

See Fig. 2.16, in which M is the larger magnitude and m the smaller. The Principle of Continuity is used in method of exhaustion, discussed next.

Method of Exhaustion

The method of exhaustion circumvents the difficulties with infinitesmals and infinite processes pointed out by Zeno (Section 2.3). It accomplishes this by replacing actual infinities by potential infinities. For example, Euclid wants to prove that the areas of circles are to one another as the squares of their diameters. He has already proved this theorem for regular polygons, so he would like to make use of Antiphon's insight that a circle can be thought of as a circle with an infinite number of sides (Fig. 2.17). In modern terms, he would like to "take the limit" and let the number of the polygon's sides go to infinity. But, instead of depending on the problematic notion of an infinite-sided polygon, Euclid applies the principle of continuity, and shows

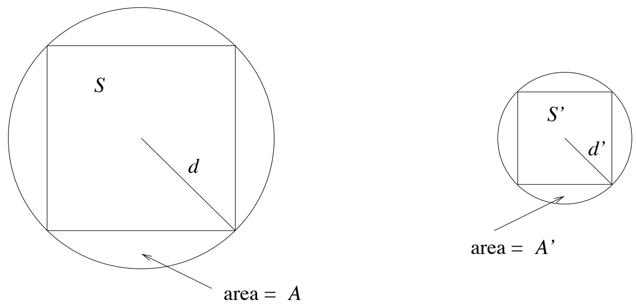


Figure 2.18: Example Application of Method of Exhaustion

that the difference between the circle and polygon can be made smaller than any given magnitude by increasing the number of sides sufficiently. This allows him to show that a contradiction would result from the assumption that the area is different from that given by the ratio of the squares of the diameters.

Application of Method of Exhaustion

The method is basically this (Fig. 2.18). Let A be the area of the larger circle. Contrary to the theorem, assume the area given by the ratio of the squares is $B < A$, that is, $A'(dd/d'd') = B < A$. Inscribe a polygon with sufficient sides so that its area is $S > B$. A similar polygon S' is constructed in the smaller circle. But since the areas of the polygons are as the squares of the diameters, it can be shown that the area of the larger polygon is less than B. Specifically, $S/S' = dd/d'd' = B/A'$. Hence $S/B = S'/A'$. But $S' < A'$, so $S < B$, which contradicts the fact that it was constructed with area greater than B. A contradiction similarly follows from the assumption $B > A$. Hence $B = A$.

2.6.4 Summary

Euclid made the deductive structure of mathematics explicit through the methods of Aristotle. However, the inability of Greek mathematics to reconcile the rational and irrational forced him to treat continuous and discrete quantities separately. In particular, he was not able to rationalize the continuous by arithmetizing geometry. The continuous and discrete remained unreconciled for over 2000 years. The arithmetization of geometry was finally accomplished around the turn of the twentieth century (see Chapter 5).

Chapter 3

Words and Images

3.1 Hellenistic Logic

> There is a certain head, and that head you have not. Now this being so, there is a head which you have not, therefore you are without a head.
>
> If anyone is in Megara, he is not in Athens: now there is a man in Megara, therefore there is not a man in Athens.
>
> If you say something, it passes through your lips: now you say wagon, consequently a wagon passes through your lips.
>
> If you have never lost something, you have it still; but you have never lost horns, *ergo* you have horns.
>
> — Chrysippus of Soli (Diogenes Laertius, 7.186–187)

> A man says he is lying. Is what he says true or false?
>
> — Chrysippus (Cicero, *De divinatione*, 2.108; cf. *Academica*, 2.96)

3.1.1 Modal Logic

During the Hellenistic period (third to first centuries BCE), significant progress was made in made in many areas of logic.[1] For example, Aristotle had begun a study of *modal* propositions (those involving possibility, impossibility, contingency and necessity), in which he enumerated the 112 possible modal syllogisms and determined which are valid (*Pr. An.* 1.3.8–22). A different system of modal logic was developed by Theophrastus (c. 370–c. 288 BCE), his successor as head of the Peripatetic school. In the next century the Megarians and the Stoics made further advances and investigated *temporal* or *tense* logic, which deals with propositions that may be true at some times but not others. Modal and temporal logics are important components of modern artificial intelligence systems (Sowa, *CS*, pp. 173–187). Carneades (214–129 BCE), the fourth head of Plato's Academy, developed a theory of qualitative probability, in which propositions could have three grades of probability: (1) "convincing," i.e., merely probable, (2) "undiverted," that is, exhibiting consistency among multiple observations, or (3) "thoroughly explored," that is, systematically tested for consistency with the rest of experience.[2] Although this is a departure from the absolutism of Plato and reflects the sceptical turn of the New Academy, it is quite consistent with a modern scientific view of truth.

3.1.2 Propositional Logic

Aristotle's logic is a logic of terms, or as we might call it now a *class logic*, a precursor to set theory. That is, it deals with logical relations between terms, such as 'man' and 'mortal', that denote classes (*eidê, forms*, in their terminology). It is generally considered that a major advance was made over Aristotle's logic when, in the late nineteenth century, logicians shifted from a logic of classes to a logic of *propositions*. In a propositional logic the variables refer to propositions (expressions that are either true or false) rather than to classes. For example, the deduction

> If it is day, then it is light;
> it is not light;

[1]Sources for this section are Long & Sedley (1987, §§ 27, 69), Bocheński (*HFL*, Pts. 2, 3), Kneale & Kneale (*DL*, Chs. 2, 3) and Hamlyn (HoE).

[2]See Sextus Empiricus *Adversus mathemeticos* (*Against the Professors*) 7.176–84, reprinted in Long & Sedley (1987, pp. 452–453).

therefore it is not day.

is an example of the schema:

If A then C;
not C;
therefore not A.

in which $A =$ the proposition 'it is day' and $C =$ 'it is light'.

Although propositional logic was rediscovered in the middle ages, Theophrastus developed the first propositional logic, and the subject was well explored by the Megarian logicians Diodorus Cronus and Philo (fourth century BCE), who developed a modern, truth-functional definition of implication,[3] and by the Stoic Chrysippus of Soli (c. 279–206 BCE), who defined schemata for a form of the "natural deduction" used by many AI systems (e.g. Cohen & Feigenbaum, *HoAI 3*, pp. 94–95); one of his rules was:

If *the first* then *the second*;
but not *the second*;
therefore not *the first*.

Here the italicized words are in effect propositional variables. In modern symbolic notation we would write:

$$F \Rightarrow S,\ \neg S \quad \vdash \quad \neg F,$$

but the difference is small. This was a further step towards reducing reasoning to formal manipulation. By all accounts, Chrysippus was one of the greatest of Greek logicians; Diogenes said (7.181–182), "if the gods took to logic, they would adopt no other system than that of Chrysippus," but none of his 705 books have survived.

[3]A truth-functional definition deals only with the truth or falsity of the propositions, not with their meaning. For example, the (material) implication (or *conditional*) 'if A then C' is considered false if A is true and C false, but true in all other cases (Sextus Empiricus, *Outlines of Pyrrhonism* 2.104–106). This definition still puzzles beginning logic students when they are told, for example, that "If pigs can fly then pigs like mud" is a true implication. The definition of implication was much debated in antiquity; Callimachus (2nd cent. BCE) said, "The very crows on the roofs caw about the nature of implications." The debate continues to this day, e.g., Appiah (*A&C*), Harper & al. (*Ifs*), Jackson (*Cond*) and Traugott & al. (*OC*).

Chysippus or some other Stoic philosopher also began the study of *semiotics* (the scientific investigation of signs in the most general sense) by carefully distinguishing three aspects of a sign. First we have the sign itself (*sēmainon*), considered as a physical phenomenon (e.g., a sound or a written text). Take as an example the name 'Socrates'. Second we have the existing object to which the name refers, for example, Socrates the person. Finally, we have the nonphysical *lekton* or "significate" (*sēmainomenon*) of the sign, which is its abstract meaning. This is meant as the objective sense of the sign (the *conceptus objectivus*), rather than any subjective impression it may cause in our mind (the *conceptus subjectivus*), which is a fourth correlate of the sign (Bocheński, *HFL*, pp. 110–111). Even though some signs, such as 'Pegasus', refer to no existing physical object, they nevertheless signify a *lekton*, since their meaning is quite definite. We know many *facts* about Pegasus. For example, it's true that he has wings, but false that he has scales; indeed, one of the facts we know is that Pegagsus doesn't exist. Logic is taken to be the science of *lekta*.

In the *lekton* theory we see again the assumption that if we can use a word meaningfully, then it must name *something*; since we can obviously talk about things that don't exist, the *lekton* ("thing said") is postulated as a surrogate object of reference. A better solution is to abandon the referential assumption, and with it the *denotational* theory of meaning, which is the more modern approach.

3.1.3 Logical Paradoxes

Beginning at least with Eubulides (fourth century BCE) the Megarian and Stoic philosophers showed great interest in logical paradoxes; a few are shown on page 61. Although these may seem ridiculous, they are not jokes; they embody serious logical issues, some of which occupied logicians for the following two millennia. The fallacies they exemplify are still pitfalls for the designers of the knowledge representation languages used in AI (see Section 3.2 for examples). The last paradox, which is the famous *Liar*, is the root of Russell's Paradox, which destroyed Frege's mathematical system, and is the inspiration for Gödel's proof of his famous Incompleteness Theorem, and for Turing's proof of the uncomputability of some functions. These topics are discussed in detail in Chapters 6 and 7. The problem continues to exercise philosophers and logicians; see for example Martin (*PL*) and Barwise &

Etchemendy (*Liar*).[4] After Galen (129–c. 199 CE), who was an excellent logician as well as a physician, logic died as a creative activity; there were only rehashes of prior work until the eleventh-century resurrection of Europe's intellectual life allowed the efflorescence of scholastic logic.

3.2 Medieval Logic

> [O]ne ought not to postulate many items when one can get by with fewer.
>
> — Ockham (Loux, *OTT*, § 12)

> Ockham's account of truth conditions will undoubtedly strike many as surprisingly contemporary, though perhaps it is fairer to say that contemporary theories should strike the reader of Ockham as surprisingly medieval.
>
> — Freddoso & Schuurman (*OTP*, preface)

> What's a' your jargon o' your schools,
> Your Latin names for horns and stools;
> If honest Nature made you fools,
> What sairs your grammars?
>
> — Robert Burns (*First Epistle to John Lapraik*)

> Unlearn'd, he knew no schoolman's subtle art,
> No language, but the language of the heart.

[4]The paradox is put in its ancient and medieval contexts in Kneale & Kneale (*DL*, pp. 16, 113–115, 227–229); van Heijenoort (LP) is a good overview of logical paradoxes. The Liar Paradox is commonly attributed to Epimenides, but was more likely invented by Eubulides, a fourth century BCE philosopher of Megara and teacher of Demosthenes. Although the problem is alluded to by Aristotle (*Sophistic Refutations* 25.180b2–7), the earliest clear statement is probably by Cicero (106–43 BCE), who attributes it to Chrysippus: "If you say that you are lying, and say it truly, you are lying" (Cicero, *Academica* 2.95). Eubulides probably used the paradox as an argument against the Platonic / Aristotelian correspondence theory of truth.

> By nature honest, by experience wise,
> Healthy by temp'rance, and by exercise.
>
> — Alexander Pope (*Epistle to Dr. Arbuthnot*, l. 398)

3.2.1 Debate about Universals

History

Medieval scholastic logic is generally considered the epitome of pedantry, yet it is relevant to contemporary approaches to knowledge representation — which are "surprisingly medieval."[5]

Issue of Objectivity

The medieval debate about the nature of universals stemmed from some problems identified by Porphyry of Tyre (232–304 CE), a Neoplatonic philosopher; among other surviving works, we have his *Life of Pythagoras*. Although the problem of universals had important implications for medieval theology, which are no longer very interesting to us, the question is still relevant to contemporary epistemology, philosophy of science and cognitive science: To what extent and in what ways are categories — in reality? — in the mind? — in language? As Porphyry put it (*Isagoge* 4.1; Kneale & Kneale, *DL*, p. 196):

> As for genus and species, I beg to be excused from discussing at present the question whether they exist in reality or have their place simply and soley in thoughts, and if they exist, whether they are separable or exist only in sensible things and dependent upon them. For such a study is very deep and requires another and larger inquiry.

In the middle ages, there were three major positions on the question, known as *realism*, *conceptualism* and *nominalism*, although we will see that the differences were not so extreme as the names suggest.

Realism

Realism held that universals exist in reality; it came in two varieties, *Neoplatonic* (e.g., Augustine, 354–430), which placed the universals in another realm (often the mind of God), and Aristotelian (e.g., Aquinas, 1225–1274), which placed them in objective similarities in particular things. In either

[5]Secondary sources for this section are Bocheński (*HFL*, Pt. 3), Kneale & Kneale (*DL*, Ch. 4) and Hamlyn (HoE).

case concepts in the mind are a result of it apprehending the *forms* in reality.

Abélard: Conceptualism

Pierre Abélard (1079–1142)[6] proposed his conceptualist theory as a third alternative to realism and the extreme form of nominalism current at that time (Roscelin's, below). He took universals to be concepts (*sermones*) in the mind, which were a kind of generic image representing the common features from the images of many particular things. When we think in terms of universals, it is these generic images that our mind manipulates. Of course, generic images would not work as representation unless there were in fact similarities between existing things, and thus in reality. So the difference between Aquinas' and Abélard's positions is more of emphasis than of kind. The notion of a generic image is very compatible with connectionist knowledge representation (discussed in Vol. 2).

Ockham: Nominalism

We find the most extreme form of nominalism in Roscelin of Compiègne (c. 1050–c. 1120), who held the universals are mere *flatus vocis* — puffs of breath. A more defensible position was maintained by William of Ockham (c. 1285–1349), who held that a universal was a particular kind of sign that can stand for many things.[7] Ockham came from the Stoic tradition, and originally held that a universal is a *logical* construct — like the Stoic *lekton* — that holds an intermediate position between the mind and the thing signified. Later, in accord with his well-known "razor" (p. 65),[8] he simplified his

[6]Also, Peter Abailard. Credited with founding the University of Paris, he was also known for his contentiousness, and was condemned by two councils, in part because his popularity with students was considered dangerous, in part because "his thoughts seemed to lead to paganism" (McKeon, *SMP*, Vol. 1, p. 207). His *Sic et non* (*Yes and No*) listed 158 theological questions on which the church fathers disagreed, and some of his works were considered heretical and burned. He married Héloïse, the niece of Fulbert, a canon of Paris, who had Abélard castrated for rescuing Héloïse from Fulbert's mistreatment and hiding her in a monastary. *Heloise and Abelard* collects their subsequent correspondence.

[7]Although Ockham (or Occam) is one of the most famous philosophers of the middle ages, his logical, physical and philosophical works are remarkably inaccessible. McKeon (*SMP*, Vol. 2, p. 351) says that few of his works have been published since the seventeenth century, and some have not been published at all. The *Quodlibeta*, selections of which he translates in his *SMP* (Vol. 2, 366–421), had not been published since 1491.

Thus we are especially fortunate for the recent publication of an English translation of the first two parts of the *Summa logicae* (Loux, *OTT*; Freddoso & Schuurman, *OTP*), but this is less than half the whole. McKeon (*SMP*, Vol. 2, pp. 351–359) is a useful source, which also contains a valuable glossary of medieval philosophical terms. Ockham was at Oxford until he was accused of heresy and had to flee to Munich.

[8]Ockham's Razor exists in many forms, although few are found in his extant writings.

account by making the universal a *mental* sign referring directly to things; thus his position was not so far from conceptualism. The idea was that these signs serve a function in mental discourse analogous to the role of words in spoken discourse (more on this below).

Direct Intuition of Particulars

A significant difference between Ockham's nominalism and the realist and conceptualist theories was in the kind of information acquired by the mind from the senses.[9] The realist and conceptualist theories claimed that the mind only knows universals (categories), which it abstracted from the images. In modern terms, cognition operates on representations constructed from general *features* computed from sense data.

> Ockham denied all this. He held that the mind could be concerned directly with the particular by means of intuitions. Intuitive knowledge is a direct knowledge of a thing or its existence. (Hamlyn, HoE, p. 16)

In other words, the concrete image of a thing is available to cognition, and cognitive processes can operate on this image rather than on some intermediate representation constructed of abstract features. This view is quite compatible with connectionist theories of knowledge, which often use concrete rather than abstract representations, and is compatible with neuropsychological evidence, which does not support feature detectors in the brain (Pribram, *B&P*, pp. 11–16, 79–81). In this context, intuition (*intuitus*) is "that by which something is known immediately, without ratiocination" (McKeon, *SMP*, Vol. 2, pp. 466–467) — that is, *nonrational, immediate cognition*.

Perfect vs. Imperfect Intuitions

However, the directness of this intuitive awareness does not guarantee its accuracy, which varies between clear and confused. Ockham made a useful distinction, between *perfect intuitions*, which are of just the present moment, and *imperfect intuitions*, which are colored by prior experience. In other words, we may have an immediate awareness of what we are experiencing right now; regardless of whether what we perceive is real or illusory, it is an accurate image of our sensations.

In the *Quodlibeta* (5, q.1) we find "*pluralitas non est ponenda sine necessitate*" — "a plurality must not be asserted without necessity." However, the commonly quoted form, "*Entia non sunt multiplicanda praeter necessitatem*" (Entities are not to be multipled beyond necessity), does not occur in his extant works. Ockham's Razor is often claimed as the basis of the scientific preference for the simpler theory (e.g., Beveridge, *ASI*, p. 116; Bronowski, *CSS*, p. 131; Joseph, *IL*, p 506).

[9] On this topic see Ockham (*Quodlibeta* 1.13, 1.15, 5.5, 6.6; McKeon, *SMP*, Vol. 2, pp. 360–375).

However, when our awareness includes past experience, then inaccuracy may enter in two different ways. On one hand we may remember a prior intuition, but such recall will likely be imperfect, since it is difficult to accurately remember a (continuous) sensory image. On the other, we are better able to remember an experience if we have grasped it intellectually, since then we have analyzed it into discrete features. But this analysis is inevitably an approximation, and so an intuition incorporating it will be imperfect. The conclusion is that prior experience imposes a bias (for good or ill) on intuitive awareness.[10] This will become a recurring epistemological theme.

Objectivity of Categories

In summary, the medieval debate over universals provides three different accounts of the objectivity of categories. Realism says they are objective because they exist in reality, either in particular things or in the mind of God; conceptualism says they are in the mind, and in this sense subjective, but get their objectivity from actual similarities of things; nominalism says that a universal is a sign that can signify any one of many particulars, and that a kind of objectivity may result either from the natural similarities of the things or from the conventions of the language.

3.2.2 Language of Logic

Logic in Context of Natural Language

Nowadays we think of logic and mathematics as closely related disciplines (indeed one is often founded on the other), but this close association did not develop before the nineteenth century (Section 4.4). Before that, and going back to Aristotle, logic was considered one of the language arts, since its main application was in judging the validity of verbal arguments. As we have already seen (Fig. 2.8, p. 33), in the middle ages logic (under the name *dialectic*) was one of the three Language Arts (the Trivium), not one of the four Mathematical Arts (the Quadrivium).

Curing the Deficiencies of Natural Language

Given that logic was investigated in the context of natural languages, it is not surprising that logicians became aware of the deficiencies of natural language as a logical instrument and tried to remedy them. This took two forms. One was to refine the language of science of that time — Scholastic Latin — into a tool for logical reasoning. We will look at this effort in this section, since it bears on some important issues in cognitive science.

Instead of refining natural languages, the second approach was to create artificial languages that were designed for logic or philosophical discourse.

[10]See Ockham (*Quodlibeta* 1.13, 1.15, 5.5, 6.6; McKeon, *SMP*, Vol. 2, pp. 360–375).

This path — which goes through Lull, Wilkins, Leibniz and Boole — leads directly to symbolic logic and AI knowledge representation languages, so it will constitute the bulk of this chapter.

What Makes a Language Ideal?

There are several assumptions underlying both of these approaches to an ideal language; we should be aware of them. First there is the realist assumption (coming from Plato and Aristotle) that the world has a definite categorical structure, and that a language is better to the extent that is truly reflects that structure. To illustrate the alternatives, we may take at the other extreme the Sapir-Whorf hypothesis (see also p. 109), which says that the structure of our language determines the structure of our reality. The view that an ideal language should be an accurate picture of reality was an essential part of the logical positivist movement in the twentieth century (Ch. 8).

Language of Thought

The second assumption — which goes back at least to Plato (e.g., *Sophist* 263e) and is still sufficiently compelling to be considered by some "the only game in town" (Fodor, *LT*) — is that there is a "language of thought." Although I will discuss criticisms of this assumption in Volume 2, here we need only observe that if there were such a language, then it would presumably be to our advantage to make our written and spoken languages as close to this as possible, since by doing so we would have eliminated all the complexities not relevant to reasoning. For this reason we have, in 1660, the famous "Port Royal Grammar" (*Grammaire générale et raisonnée* by Arnauld and Launcelot), which sought "to set forth what is common to all languages" and "to explain language by reference to the constitution of the human mind" (Ellegård, 1973, pp. 667b–668a); it was followed shortly by the influential "Port Royal Logic" (*Logique ou l'art de penser*, Arnauld & Nicole, 1662). But they are just the tip of the iceberg.

The two assumptions — a categorical structure in reality and a language of thought — together generate a pervasive, rationalistic optimism that many of our problems would be solved if we could simply use the correct language to talk about them. Thus the quest for an "ideal" language.

* * *

Ockham: Accomplishments

In presenting some relevant developments in medieval logic, I will be drawing again from the works of William of Ockham.[11] We are using but

[11] Sources are Parts 1 and 2 of Ockham's *Summa logicae* (Loux, *OTT*; Freddoso & Schuurman, *OTP*) and the *Quodlibeta* (McKeon, *SMP*, Vol. 2, pp. 360–421).

a fraction of his research in meaning and reference, temporal and modal logic, nominalism, and many other topics. Here we consider his theories of *intention* and *supposition*.

Intentionality

The nature of *intentionality* may be the central philosophical problem in cognitive science and artificial intelligence. Although it originated in the middle ages, it is still widely debated (e.g., Dennett, *IS*; Diamond & Teichman, *I&I*; Dreyfus & Harrison, *HICS*; Searle, *Int*), and many philosophers believe that intentionality is a characteristic of natural intelligence that is beyond the reach of artificial intelligence.[12] In considering this question we must be careful to distinguish the technical term *intention* from its everyday meaning, "a goal, aim or objective."

Intentionality is that property of a mental state by which it is *about*, or *directed at*, something else. It derives from Latin *intendo*, which means to stretch toward, to point at, or to direct one's mind toward; and at least from Cicero's time, the noun *intentio* could refer to the acts of stretching, reaching, and concentrating one's attention (Glare, *OLD*, s.vv. intendo, intentio). The Schoolmen adopted it as a technical term for the ablity and act of the mind pointing outside of itself (McKeon, *SMP*, Vol. 2, pp. 465–466).

Ockham's Definition

Since Ockham views thought as "mental discourse," he treats an intention as a word in the mental language:

> an intention of the soul is something in the soul capable of signifying something else... (*Sum. log.* 1.12; Loux, *OTT*, p. 73)

More specifically, an intention "is either a sign naturally signifying something else (for which it can supposit) or a potential element in a mental proposition" (*Sum. log.* 1.12; Loux, *OTT*, p. 74). ('Supposit' is a scholastic technical term meaning 'stand in place of'; we'll encounter it again.) Thus for Ockham intentionality is akin to the referential relation between a sign and the thing it signifies.

First Intention

Avicenna (Ibn Sīnā, Abū' Alī al-Husayn, 980–1037) had distinguished two kinds of intentions.[13] In Ockham's formulation, a *first intention* is a mental sign that does *not* signify an intention or sign. Thus most everyday concepts, such as *man*, *blue*, and *fire*, are first intentions.

Second Intention

[12] Gregory (*OCM*, s.v. Intentionality) is a convenient overview.

[13] In translating Avicenna, *intentio* was apparently chosen as a translation of the Arabic *ma'nā*, a meaning, thought, signification or notion (Butterworth, *AMC*, p. 25, n. 2; Kneale & Kneale, *DL*, p. 229).

A *second intention*, in contrast, is a mental sign of a first intention; examples are logical terms such as *species* and *genus*.

That is, a first intention, such as *man* refers to things outside the mind (e.g., Socrates, Hypatia and other people), whereas a second intention, such as *concept*, refers to things inside the mind (e.g., *man*, *blue* and other categories). Ockham claims that the terms of logic are of the second intention (because logic studies first intentions), whereas the terms of the other sciences are of the first intention (because they study things other than intentions).

Relevance of Supposition

The theory of *supposition* may seem a clear example of Scholastic pedantry, but it's an important knowledge representation issue in AI. Consider the following true statements:

Man is mortal.
Man is a general term.
Man means an adult human male.
Man is an English word.
Man is a three-letter word.
Man is a one-syllable word.
Man rhymes with can.
Man is in italic type.
Man is the first word in this line.
Man is the tenth occurrence of Man in this display.

In each of these true statements 'Man' is used in a different way and in fact refers to something different. It is worthwhile to go over these sentences again and be sure that in each case you can say what 'Man' denotes. To help, try replacing 'Man' by other expressions, such as 'Homo sapiens', 'Mensch', 'Mann',[14] 'Dan', 'Plan' and '*anthrōpos*', and make a table of the truth or falsity of each proposition. Notice that normal quotation marks are not sufficient to eliminate the difficulties; try substituting ' "Man" ' and ' "Plan" ' for 'Man'.

This is an important issue for knowledge representation languages, since we do not want to allow inferences such as this:

$$\frac{\begin{array}{r}\text{Man is a three-letter word;}\\ \text{Dan is a man;}\end{array}}{\text{therefore, Dan is a three-letter word.}}$$

[14] In German, 'Mensch' generally means any human being whereas 'Mann' refers to a human male.

which happens to reach a correct conclusion, or this (cf. Sowa, *CS*, p. 84):

$$\frac{\begin{array}{r}\text{Elephant is a general term;}\\ \text{Jumbo is an elephant;}\end{array}}{\text{therefore, Jumbo is a general term.}}$$

which reaches a incorrect conclusion. Yet this is exactly what will happen if they blindly implement a simple deduction rule such as:

$$\frac{\begin{array}{r}M \text{ is } P;\\ S \text{ is } M;\end{array}}{\text{therefore, } S \text{ is } P.}$$

We recognize the absurdity of these inferences, but a rule-based expert system does not have an experiential basis for recognizing sense and nonsense; therefore it's necessary to block inferences of this kind by syntactic mechanisms (such as quotation marks) or other formal devices.

Supposition Defined

These examples illustrate that there are many ways that a word can *supposit*, or stand, for other things. In 'Man is mortal', the word 'man' stands for any human being; in 'Man is a general term' it stands for the concept or category *man* = *human being*; in 'Man is an English word' it may stand for either the written or spoken word 'man'. The Schoolmen called this property of standing for something else *supposition*: "Supposition is said to be a sort of taking the place of something else." (Ockham, *Sum. log.* 1.63; Loux, *OTT*, p. 189) Notice that supposition is a property of a term *in the context of a proposition*; that is, 'Man' supposits in a different way in each of the examples above. Ockham distinguished three kinds of supposition (although he also recognized finer distinctions could be made).

Personal Supposition

First, Ockham says *personal supposition*:

> occurs when a term supposits for the thing it signifies, whether this thing be an entity outside of the soul, a spoken word, or any other thing imaginable. (*Sum. log.* 1.64; Loux, *OTT*, p. 190)

An example of personal supposition is 'man' in 'man is mortal', since in the proposition 'man' supposits for (stands in place of) Socrates, Plato, Hypatia, or any other person.

Simple Supposition

Next he says, "Simple supposition occurs when a term supposits for an intention of the soul and is not functioning significatively" (*Sum. log.* 1.64; Loux, *OTT*, p. 190). For example, in 'Man is a species' the term 'man' does not stand for individual persons, as in the previous example, but instead for the concept (first intention) *man*. *Man* is a species, but Socrates and Hypatia are not; conversely, the category *man* is not mortal but Socrates and Hypatia are. Ockham explains that a term in simple supposition is "not functioning significatively" because 'man', for example, signifies persons not intentions; as we saw it is a term of first intention, not second intention.

Therefore, when 'man' is in simple supposition, it indicates the intention *man* indirectly. That is, a term in simple supposition stands for an intention in much the same way a part can refer to a whole by *metonymy*, for example, when we say "I don't have any wheels" to mean "I don't have a car." In everyday speech we rely on context to distinguish personal and simple supposition; in written language we might use italics to signal simple supposition, as I have done in this paragraph.

Material Supposition

Finally, "Material supposition occurs when a term does not supposit significatively, but supposits for a spoken word or a written word" (*Sum. log.* 1.64; Loux, *OTT*, p. 191). For example, in the proposition "'Man' is a name", the word 'man' refers neither to people nor to concepts, but to certain physical phenomena, either vibrations in the air (spoken words) or visible marks on paper (written words). Once again, in spoken language, context tells us when material supposition is intended; in written language we often use quotation marks, as in this paragraph.[15]

Kinds of Material Supposition

As the preceding examples demonstrate, there are actually several kinds of material supposition. For example, 'man' may refer to the spoken word (as when we say it has one syllable), or to the written word (as when we say it has three letters), or to a particular printed form (as when we say it is in italics), or to a particular instance of that form (as when we say it is the tenth occurrence on the page). For everyday purposes these are hairs that need not be split, but they can trip up formal deductive systems.

Use vs. Mention

Modern philosophy simply distinguishes between the *use* and the *mention* of an expression. Personal supposition *uses* the term (significatively); simple and material supposition *mention* the term, either to refer indirectly to a mental sign (simple supposition) or to a spoken or written sign (material supposition).[16]

Object vs. Metalanguage

[15]Whenever there is a chance of confusion in this book, I will follow the convention of using italics for simple supposition and single quotes for material supposition. For example: Man is mortal; *Man* is a species; but 'Man' is an English word. For other purposes, I use double quotes; as usual italics are also used for emphasis.

[16] Alternately, modern philosophy distinguishes *reference* — or extension — from *mean-*

In the theory of supposition we see the fourteenth century logicians' version of the object-language/metalanguage distinction, which is an important tool of modern logic. A *metalanguage* is used to talk about another language, the *object-language*. In fact the two may be the same language, as when we use English to talk about English. As we've seen, everyday language rarely makes these distinctions, but many philosophers think that a confusion of metalanguage and object-language is the root of many paradoxes and fallacies; we'll see an example shortly. The distinction is also fundamental to the proofs of Gödel's incompleteness theorem, Turing's incomputability theorem, and many other important results (Chapter 7).

Sophismata

During the thirteenth and fourteenth centuries the study of paradoxes (*sophismata*) provided a vehicle for studying many important problems in meaning and reference.[17] For example Jean Buridan (c. 1295 – after 1358) proposed paradoxes of this kind: I show you my hand and ask if you know whether the number of coins in it is odd or even. Of course you say you do not know. But then I open it and show three coins, and say, "You claimed that you didn't know if the number of coins is odd or even, but the number of coins is three, therefore you have said that you don't know whether three is odd or even." The number of coins equals three, and I have simply substituted equals for equals, 'three' for 'the number of coins', in your assertion. Where is the fallacy?

Referential Opacity

In modern terms, we say that *intentional contexts are referentially opaque*, but this requires some explanation. An intentional context is an intention or a description of an intention, and so it is *about* something else. In particular, verbs of knowing, believing, hoping, fearing, etc. describe intentions, because you know *that* something or believe *that* something, etc. An expression is in a *referentially transparent* context when the expression can be replaced by any other with the same referent without changing the truth of the proposition; it is referentially opaque when this is not the case, that is, when equals cannot be substituted for equals. To return to Buridan's example, 'the number of coins is odd' is referentially transparent since its truth value doesn't change when 'the number of coins' is replaced by any other expression with the same referent. Since 'the number of coins' and 'the square root of nine' both have

---

ing — or intension. That is, Plato, Hypatia, etc. are the *referents* of 'man', and the class of all people is the term's *extension*; but the concept or intention *man* is the meaning of 'man', and the properties of human beings comprise the *intension* of the term; extension and intension are explained later (Section 4.3.2).

[17]See Moody (ML); Bocheński (*HFL*, § 35) contains a selection of medieval sophismata.

the same referent, three, one may be substituted for the other in this or in any referentially transparent context. For example, from three being the square root of nine we could conclude that the number of coins is the square root of nine.

The designers of knowledge representation and inference systems must be careful to distinguish transparent and opaque contexts; indeed, it is even important in reasoning about programs (MacLennan, *FP*, pp. 11–13, 100–101). Further, referential opacity is often considered a hallmark of intentional and mental states and events (Dennett, *IS*, pp. 174–175, 184–185, 240–242).

The Liar Paradox

Of course, the Schoolmen were especially interested in The Liar, and it kept them busy for two centuries. Many books were published with names such as *The Unsolvable* (*Insolubile*), which remind us of twentieth century works such as Gödel's "Undecidable Propositions…". However, there was no shortage of "solutions"; by 1429 Paul of Venice had published fourteen solutions from his predecessors and added a fifteenth of his own (*Logica magna*; Bocheński, *HFL*, pp. 241–251).

Ockham's solution

Ockham was one of the first to treat The Liar as a serious logical problem, and a chapter of his *Summa logicae* (Part 3) was called "About Insolubles." His solution was to claim that the self-contradictory statement was illegitimate, because the subject of the proposition was the proposition itself; thus it was being simultaneously used and mentioned. This solution was not considered adequate at the time, and in fact did not handle paradoxes such as:

> Socrates believes this and no other: 'Plato is deceived' … but Plato … believes this: 'Socrates is not deceived'. (Bocheński, *HFL*, 35.14)

Here we have a reciprocal reference between two propositions: S affirms P, but P affirms not-S, yet neither directly mentions itself.[18]

Buridan's Solution

Buridan devised a better solution, which is quite contemporary in approach. He observed that the semantics of the language creates a correspondence between each object-language proposition P and a corresponding metalanguage statement, "'P' is true". Therefore, the following two propositions must have the same truth value:

[18] In computer science this is called *indirect recursion*, as opposed to the *direct recursion* of The Liar.

$$\text{this is false.} \iff \text{‘this is false’ is true.}$$

The result is a contradiction. For if Q = ‘this is false’, then the meaning of Q is that Q is false. But Q also implies the right-hand side, which says Q is true. Since Q implies contradictory propositions, it must in fact be false. As Moody (ML, p. 534) puts it, “a proposition depends on the semantical structure of the language to which it belongs, so that it cannot be used to violate the conditions which give it its status as a linguistic expression.” Investigation of The Liar continues still (e.g., Barwise & Etchemendy, *Liar*; Martin, *PL*), and it is the basis of Gödel’s and Turing’s proofs of the limitations of formal logic and digital computation (Ch. 7).

Summary

We’ve seen that in the course of trying to use Latin as a logical language, the Schoolmen investigated a variety of deep semantical problems in logic and language, and developed sophisticated solutions to many of them, including the theories of intention and supposition. On the other hand, the very difficulty of working with a natural language may have opened the way for the creation of ideal, artificial languages for logic and science. In Section 3.4 we’ll look at Lull’s early, naive, but very influential approach, and later consider the much more sophisticated systems of Wilkins, Leibniz and Boole. Out of them symbolic logic developed into a powerful tool for knowledge representation and inference, which encouraged its later use in AI and cognitive science.

3.3 Combining Images and Letters

3.3.1 The Art of Memory

Simonides of Ceos: 556–468 BCE

Before addressing Lull and his Art, I consider briefly several ideas that lurk in the background of the theories of thought and computation that are the topic of this chapter. We begin with the Art of Memory (*Ars Memorativa*), which has its roots in ancient Greece, where the poet Simonides of Ceos was credited with its invention.[19] It was practiced, especially by orators, throughout antiquity and continued to be popular through the middle ages and Renaissance; indeed the techniques are still taught (e.g., Bellezza, *IYMS*).

The Method of Places

[19] Principal sources for this subsection are Yates (*AoM*), Bolzoni (*GM*), and Small (*WTM*, Pt. II). The latter deals with the cognitive science of memory techniques, as does McDaniel & Pressley (*IRMP*).

One of the central techniques (credited to Simonides) was called the *method of places* and was used often to remember speeches. First, each topic in the speech was associated with a vivid, emotion-laden and therefore memorable *image*; in this way the individual topics were remembered. Active images (*imagines agentes*), incorporating motion, were especially recommended. Then the information was organized — the topics placed in the correct order — by making use of human spatial memory, for the ancient technique was to imagine some familiar place, such as the forum, with a number of distinct *places* (Grk. *topoi*, Lat. *loci*), for example, temples, fountains, shops, and street corners. In the imagination, the images were put in the selected places in the order that the places would be reached in a walk, and later the speaker remembered the topics in the correct order by imaging a walk from each place to the next. At each place, the image reminded the speaker of the topic to be addressed at that point in the speech.[20]

Summary

In summary, the method of places made use of emotion-laden sensuous images located in physically real places, with the spatial organization of the places representing the logical organization of the individual topics. In modern terms the image is an *encoding* technique that gives a distinctive visual representation to an abstract topic, and the places constitute a unified spatial representation for the topics' *organization*; they correspond to visual-processing pathways in the brain specialized to *what* and *where*.

Later Developments

As the art of memory developed over the centuries, new techniques were developed as well as new applications. One new development was the use of "fictitious places," such as imaginary palaces or theaters, which could represent abstract and complex relationships among the memorized ideas. There were even attempts, such as Giulio Camillo's (c.1480–1544) Memory Theater, to construct physical models of these memory structures, in which physical images of important ideas were arranged in an architectural space. In the Middle Ages the art of memory was used for devotional practices, such as contemplating the virtues and vices, images of which were organized in an imaginary physical structure, such as a tree, tower, or ladder. In this way the ideas were impressed on the memory, which was supposed improve one's character (a purpose of the art of memory since ancient times). In the Renaissance and Baroque periods, the art of memory was explored as a

[20] Yates (*AoM*, p. 46) remarks that the widespread use of the method of places is probably the reason that we use the word *topic* for the subjects located at the places (*topoi*). "Common places" (Lat., *communes loci*; Grk. *koinoi topoi*) were used for typical topics, which are still called "commonplaces" (Small, *WTM*, p. 90).

method for organizing all human knowledge, as will be discussed later.

3.3.2 Combinatorial Inference

Combinatorial Approach to Inference

Lull's Ars Magna takes a *combinatorial* approach to inference, so it is worthwhile to consider this before looking at his system. Recall that Aristotle classified all syllogisms into three *figures* (Section 2.5.4). Furthermore, he recognized propositions in four forms, which in the middle ages were called A, E, I and O:

A:	Universal Affirmative
E:	Universal Negative
I:	Particular Affirmative
O:	Particular Negative

The A and E forms are called *universal* propositions, the I and O *particular*.[21] For example,

A	All men are mortal
E	No insects are mammals
I	Some men are Greek
O	Some men are not Greek

In principle, each of the three propositions that constitute a syllogism could be in any of these four forms. Thus we have the $4^3 = 64$ possible *moods* of the syllogism: AAA, AAE, ..., OOI, OOO. Since each mood can occur in three figures, there are a total of 192 possible Aristotelian syllogisms.

Generate and Test Procedures

Certainly not all of the 192 possible Aristotelian syllogisms are valid, and Aristotle had already determined which are valid and which aren't. In the process he had identified a number of general rules that must be obeyed by any valid syllogism. Later commentators listed the valid syllogisms by a process of elimination: enumerate all the possible syllogisms and then strike out those which are invalid.[22] "Generate and test" procedures of this kind are still widely used in artificial intelligence and other computer applications (see also p. 137).

The Problem of Combinatorial Explosion

[21] These abbreviations can be remembered by the mnemonics **A**ff**I**rmo (for the affirmations) and n**E**g**O** (for the denials).

[22] For example, Ockham enumerated the 1368 modal syllogisms in his formulation, and identified nearly 1000 that are valid.

The problem with *combinatorial* procedures such as these is that reasonably rich rules of combination may lead to a *combinatorial explosion* of possibilities. We have seen that even the simple Aristotelian syllogism leads to $3 \times 4 \times 4 \times 4 = 192$ possibilities. In general the number of items to be enumerated increases exponentially with the size of the items (e.g., 4^3 syllogisms composed of 3 propositions). For practical sized problems, testing all the combinations may even exceed the capabilities of modern computers. Thus combinatorial explosion remains a problem in contemporary combinatorial algorithms. It is often the reason that a demonstration AI program will not "scale up" to practical sized problems.

3.3.3 Kabbalah

Kabbalah Defined

Another source of Lull's ideas was *Kabbalah* (Hebrew for "tradition"), a kind of mystical Judaism that was becoming popular when Lull was active, and that includes a number of ideas and practices that are relevant to our topic.[23]

Importance of Hebrew Alphabet

Fundamental to the Kabbalah is the idea that the structure of the universe is reflected in the Hebrew alphabet. Therefore the truths of heaven and earth can be discovered by contemplation and manipulation of the letters. This belief is based in part on the idea that the Torah (Law), written in Hebrew, reflects the *logos* (the rational structure) of the world, that is, the realm of archetypal forms or ideas; thus kabbalists identified the Torah with wisdom and the active intellect. As we will see, kabbalistic interpretation of the Torah suggested a process for generating knowledge though the manipulation of symbols.

The Sefirot

Kabbalists explain the creation of the world as an emanation from *En Sof* (the limitless and unknowable — that is, *in(de)finite* — God) down through ten *sefirot* (Heb., spheres) of increasing degrees of determination and delimitation, terminating ultimately in the material world. These spheres correspond to the Decad, the numbers 1 – 10, an idea with Pythagorean roots

[23]Also transcribed cabala, cabbala, kabala, qabalah etc., from Hebrew *qabālāh* from קִבֵּל (receive, take); some authors use the variant spellings to distinguish the original Jewish tradition from its later use by Christians and Renaissance occultists, but there is no widespread convention. (For the most part I have used ALA-LC Romanization conventions.) The books by Gershom Scholem (e.g., *Kab*, *K&S*, *Z*) and by Moshe Idel (e.g., *K*, *AP*) are authoritative. The principal sources for this section are Eco (*SPL*, ch. 1) and Poncé (*K*).

(Scholem, *K&S*, p. 167), and are traditionally named Crown, Understanding, Wisdom, Power, Mercy, Beauty, Glory, Victory, Foundation, and Kingdom.[24] As the principal determinate emanations of God, the *sefirot* are taken to be the chief divine attributes, the primary names of God.

The Tree of Life

The *sefirot* are often displayed in a geometrical arrangement called the Tree of Life (*Otz Chiim*), in which relationships between certain of these divine attributes are represented by lines connecting them. (As will be discussed later, there is a connection here with the use of trees and other "fictitious places" in the Art of Memory.) In the most common such diagram, there are 22 connections, corresponding to the 22 letters of the Hebrew alphabet, also considered divine names (see Fig. 3.1). The *Sefer Yetsirah* (*Book of Creation*), one of the primary kabbalistic texts, says that the sefirot and the Hebrew letters together are the 32 "paths of wisdom," which are the "stones" from which God created the universe, "by number, writing, and speech" (Wescott, *SY*, 1.1). Elsewhere (2.4), it says that God affixed the 22 letters to a wheel with 231 gates, which He rotated forward and backward (a process imitated, as we'll see, by Lull); note that 231 is the number of combinations of two different Hebrew letters. In Ch. 4 the text observes that God was able to produce the inconceivable diversity of the world through the permutations of larger numbers of letters, for there are $2! = 2$ permutations of two particular letters, $3! = 6$ permutations of 3 letters, $6! = 720$ of 6 letters, $7! = 5040$ of 7 letters, and so forth. The number of permutations of the letters of the Hebrew alphabet is $22! \approx 1.124 \times 10^{21}$, an astronomical number. Thus combinatorial explosion, the generative productivity of a finite set of symbols, accounts for the limitless multiplicity of our world.

Reality Reflected in Arrangements of Letters

From this kabbalistic perspective, the Hebrew names for things are not arbitrary, but reflect in the arrangements of their letters the true nature of things. Therefore one can discover truth both by investigating the arrangements of letters in a sacred text, such as the Torah (encoding the *logos* of the world), and by contemplating new arrangements generated by combinatorial processes.

Gematria and Other Kabbalistic Practices

For example, hidden connections between words are discovered by means of *gematria*, a practice based on the assignment of numerical values to Hebrew letters. Thus the messiah (*mashiah* משיח) is related to the brass serpent (*nahash* נחש) of Moses because the two words add up to the same numerical

[24]Heb.: *Kether, Hokhmah, Binah, Gevurah, Hesed, Tifereth, Hod, Netsah, Yesod, Malkuth*, respectively. As usual, there are transcriptional variations.

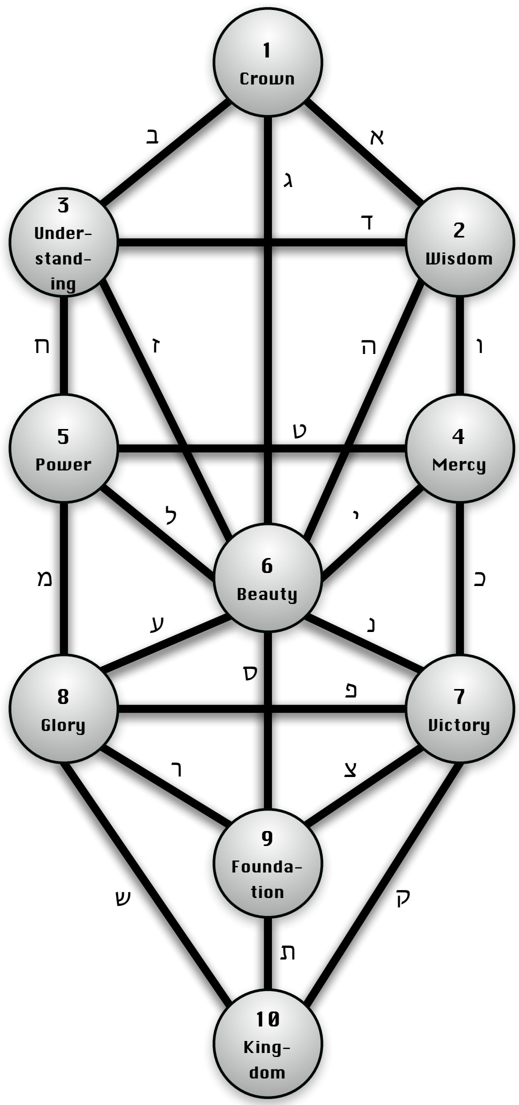


Figure 3.1: The Kabbalistic Tree of Life. The spheres (*sefirot*) represent a decad of divine attributes or "emanations" corresponding to the numbers one to ten. The 22 lines, which relate certain pairs of attributes, are marked with the 22 letters of the Hebrew alphabet. The ten sefirot and 22 connections together constitute the 32 "paths of wisdom."

value (368). As a result, underlying the surface text of the Torah, kabbalists discover hidden connections established by quantitative relationships (a conceptual precursor to the scientific identification of hidden quantitative relationships underlying perceptible reality). Other hidden meanings are discovered by other formal manipulations of the text (e.g., taking the first letters of every word in a phrase, or the last letters; exchanging corresponding letters in the first and second halves of the alphabet). Of course it was not uncommon to seek unapparent, allegorical, hidden, or esoteric meanings in sacred texts (it was also common in Christian and Islamic exegesis, for example, and even applied to the Homeric epics); what is important to notice here is the emphasis on the numerical and combinatorial manipulation of letters interpreted as archetypal ideas.

Permutation and Method

According to the rules of gematria, the sacred four-letter name of God, YHWH (יהוה), has the value 72, and indeed God is said to have 72 divine names. Contemplation of the names of God was a common practice in Christianity and Islam as well as in Judaism, but in kabbalah it took an especially combinatorial form. For example, the kabbalist Abraham Abulafia (1240–91, contemporary with Lull) described a contemplative practice in which one turned the divine name "like a wheel," generating various permutations, until it produced a word of wisdom. In effect, he was proposing a mechanical *method* for generating knowledge, a goal that also lies behind the attempts to mechanize thought, which we will discuss, and the development of scientific method. These techniques were supposed to work because, it will be recalled, the letters were the constituents from which God created the world.

3.4 Lull: Mechanical Reasoning

> The understanding longs and strives for a universal science of all sciences, with universal principles in which the principles of the other, more special sciences would be implicit and contained as is the particular in the universal ...
>
> — Ramon Lull (*Ars magna et ultima*, 218; Bocheński, *HFL*, § 38.01)

> The subject of the Art is the answering of all questions, assuming that

one can identify them by name.

— Ramon Lull (*Ars brevis*, prologue; *SWRL*, p. 579)

Lullum, antequam Lullum noscas, ne despicias.

— A. Oliver (*Raymundi Lulli opera medica*)

Question: Whether God exists.
Solution:

AA	being perfection	privation imperfection	SV	YZ

— Lull (*Ars demonstrativa*; Llull, *SWRL*, p. 444)

As for Astronomy, study all the rules thereof, let passe neverthelesse, the divining and judicial Astrology, and the Art of Lullius, as being nothing else but plain abuses and vanities.

— Rabelais (*Gargantua and Pantagruel*, Bk. 2, Ch. 8)

After salutation, observing me to look earnestly upon a frame, which took up the greatest part of both the length and breadth of the room; he said, perhaps I might wonder to see him employed in a project for improving speculative knowledge by practical and mechanical operations.

— Swift (*Gulliver's Travels*, Pt. 3, Ch. 5)

3.4.1 Background

Ramon Lull: 1232–1315

Ramon Lull would likely be a minor figure in the history of philosophy were it not for his invention in 1274 of the *Ars Magna* (Great Art), which exerted a powerful influence on later philosophers, and indirectly on AI and cognitive science.[25]

He lived the life of a rake for some 30 years, until he saw a series of visions and became devout. After nine years of study, he spent a week contemplating God on Mt. Randa (near Palma) in 1274, and had "revealed" to him the Great Art; as a consequence he was later known as *Doctor Illuminatus.* After his illumination Lull became a prolific author, producing over 260 works (fiction as well as nonfiction), of which nearly 240 survive. Most of them were composed in Catalan or Arabic, and he effectively created literary Catalan.

After Lull's death, *Lullism* began to grow, slowly at first in the fourteenth and fifteenth centuries, but then flourishing in the sixteenth and the first half of the seventeenth. Here we find it exerting its influence on the coinventers of the calculus, Leibniz (Section 4.3) and Newton (who owned eight volumes of Lull, including six, spurious alchemical works[26]). It also inspired, as we'll see, the pursuit of universal languages for representing existing knowledge and of systematic methods for generating new knowledge.

Aspects of Lullism continue in many forms to this day. For example, his Tree of Science (*Arbor scientiae*) proposed a unified hierarchy of all the sciences, which anticipates the "Unity of Science Movement" in 20th century logical positivism (Ch. 8).

3.4.2 Ars Magna

Use of Combinatorial Devices

Like his predecessors, Lull takes a combinatorial approach to inference, but his Ars Magna is different from their approaches in two important respects.

First, he used simple devices for enumerating all the combinations (discussed later). Although the idea is simple, and others may have done this before him, Lull is probably the first to use mechanical aids systematically. His approach was to identify the most basic concepts in any field of study, and then to use his devices to generate all possible combinations of these concepts. He believed that by contemplating the possible relationships of

[25]There is considerable variety in the spelling of Lull's name; one finds Lull, Lully, Lulio, Lullius, Lulle and Llull, as well as Ramon, Ramón, Raimundo and Raymund. Although the modern Catalan spelling is Ramon Llull, I will use the more familiar Lull.

The primary source for this section is Llull (*SWRL*); secondary sources include Yates (*AoM*), Gardner (*LMD*, Ch. 1) and Johnston (*SLRL*); Llull (*DI*) is a convenient sample of his work.

[26]Spurious, because there is no evidence that Lull was an alchemist, per se.

these elementary concepts he could discover the primary truths of that field.

A Universal Art

Second, Lull did not restrict his Great Art to a narrow domain, such as the science of the syllogism. Rather, he viewed his art as a means of discovering the deepest truths in every field of knowledge. Thus Lull believed he had invented a device that would allow him to generate mechanically the primary truths of all the sciences. Although he was incorrect in this belief, his vision inspired many later investigators of automated reasoning (Section 4.3 and Chapter 9) and contributed to the search for a reliable *method* for discovery of knowledge. Lull's aims were similar to the kabbalists in that he thought he was exploring the archetypal ideas underlying and permeating the world.

Elementary Concepts

A premiss of Lull's approach is that there *are* elementary concepts in a given knowledge domain. That is, he was assuming a kind of *logical atomism*, the view that all categories can be analyzed into a certain number of *atomic categories* that cannot be further subdivided.[27] Lull believed that by exploring all the possible relations between the atomic categories he would discover the primary truths (like those stemming from Aristotelian definitions, p. 49). These primary truths would in turn imply other truths, including relationships between composite categories (those defined in terms of the atomic categories).

Lull's Combinatorial Devices

The simplest Lullian device is a circle circumscribed with the basic concepts of some domain. Lines connect all the possible two-term combinations (Fig. 3.2, p. 91); some of Lull's devices involved two or more concentric disks (Fig. 3.5, p. 94). By rotating these disks all the two (or more) term combinations could be generated (reminiscent of kabbalistic rotation through the permutations of the Hebrew alphabet). These devices have motivated some to call Lull the "father" of computer programming (Moody, ML, p. 530)! To see if this claim is at all justified, we will have to look at his Art in more detail.

Ars demonstrativa

In an early version of the Ars Magna (presented in *Ars demonstrativa*, c. 1283) Lull defines an "alphabet of thought," which is represented by the 23 letters of the medieval alphabet (ours without J, U, W). The 16 letters B to R represent the basic *terms* of various subjects; the remaining seven letters

[27] Leibniz pursued a more sophisticated logical atomism; see Section 4.3. Logical atomism reached the height of its popularity in the twentieth century; see Section 8.3 and Chapter 9.

(ASTVXYZ) name various *figures* used to structure the concepts represented by the terms. The subject domains represented by the figures are:

A	Attributes of God
S	Actions of the rational soul
T	Principles and relations
V	Virtues and vices
X	Predestination and objectification
Y	Truth
Z	Falsehood

Except for Y and Z, which represent truth values, and so are atomic, all the other figures contain the 16 terms B–R connected by colored lines in various patterns. In some cases Lull's figures are quite simple, only showing combinations of two terms (see Fig. 3.2, which is from a later version of the Art employing only nine terms), whereas others were much more complex.

In addition to these figures, there was a Demonstrative Figure (the most complex, with six concentric moving rings), an Elemental Figure, which displayed combinations of the four elements in a square matrix, and figures containing the 16 elementary terms of various specialized disciplines, such as philosophy, theology and the law.

The Computer Analogy

Bonner (Llull, *SWRL*, pp. 309–310) has described Lull's Art in computer terms. The figures AVXYZ, the Elemental Figure, and the figures for the special sciences all correspond to the basic data or "knowledge base" upon which the system operates. Figure T, which is a relational figure, corresponds to the processing unit, since it establishes internal relations among the terms. Figure S is also a relational figure, but it establishes external relations, either between the terms and the operator, or between the terms and the person to be convinced by the argument. (In effect it's a link between the object and meta levels.) Thus it corresponds to a control or input-output unit.

Needless to say, the computer analogy cannot be carried too far, for not only does the Art require a human to sequence the steps (as on a hand calculator), but human interpretation is also necessary for correct execution of the individual steps (as we'll see shortly). Nevertheless, by reducing a wide range of inquiries to a methodical generate-and-test process, the Art takes an important step toward the mechanization of reasoning.

Exhaustive Solution of Problems

One thing which Lull's Art permits is systematic investigation of *all* the questions in a given domain; thus its combinatorial approach aids completeness. For example, one part of the *Ars demonstrativa* is devoted to systematically posing and solving 1080 questions; in 39 cases the method of solution is explained; in the remaining 1041 Lull gives only the "compartments" that show the relation of the terms — in effect, they are left as "exercises for the reader."

Examples of the Art

To give a bit of the flavor of the Lullian Art, I will paraphrase and explain one of his "demonstrations" (Llull, *SWRL*, p. 455):

> *Question:* Which is more demonstrable, truth or falsehood?
> *Solution:*

YZ	EAVY	IVZ	EVZ	IAVY	XX

> Y [truth] accords with being and perfection, and Z [falsehood] with privation and imperfection. The first X [objectification?] of the last compartment signifies the first concordance in the second and third compartments.
>
> In the second we have EAVY, where E = B & C & D = remembering & understanding & loving, and AVY = God & virtue & truth; in the third, IVZ, where I = F & G & H = remembering & understanding & hating, and VZ = vice & falsehood.
>
> The second X of the last compartment signifies the second concordance in the fourth and fifth compartments: In the fourth, EVZ = remembering, understanding & loving both vice & falsehood; and in the fifth, IAVY = remembering, understanding & hating AVY, namely, God, virtue & truth.
>
> "This being the case, the question is therefore solved by means of the above-stated signification."

And two "exercises for the reader" (Llull, *SWRL*, pp. 476, 540):

> *Question:* Whether the virtue of the soul exists in the powers in continuous or discrete quantity.

Solution:								
FF	EI	NR	AS	SS	ST	fire fire	fire earth	fire air

> *Question:* How do angels speak to one another?

Solution:			
intelligence habit	AA	dignities act	form relation

Table 3.1: Meaning of Terms in Lull's Great Art

Terms	Absolute Principles	Relative Principles	Questions	Subjects	Virtues	Vices
B	goodness	difference	whether?	God	justice	avarice
C	greatness	concordance	what?	angel	prudence	gluttony
D	eternity	contrariety	of what?	heaven	fortitude	lust
E	power	beginning	why?	man	temperance	pride
F	wisdom	middle	how much?	imaginative	faith	apathy
G	will	end	of what kind?	sensitive	hope	envy
H	virtue	majority	when?	vegetative	charity	ire
I	truth	equality	where?	elementative	patience	lying
K	glory	minority	how?	instrumentative	pity	inconstancy

charity justice	beginning end	concordance middle

You may draw your own conclusions ...

Ars brevis

Beginning about 1290 Lull greatly simplified his system, supposedly because people found the complete system too difficult. The new version is presented in the *Ars brevis* and *Ars generalis ultima*; we'll consider it briefly to get a clearer idea of the Lullism.

Terms

The basic terms in the simplified Art are the nine letters B to K (omitting J), which represent the members of several sets of nine categories (Table 3.1). Thus, depending on context, E could be power, beginning, man, pride, etc. Since the basic terms have meanings, they cannot be considered entirely formal; nevertheless, a letter's meaning in a particular context is determined largely by its relation to the other letters in the Lullian figures. For reasons to be explained, Lull believed that these categories were fundamental to all knowledge and indeed to the structure of the world, and so he calls them reasons (*rationes*), dignities (*dignitates* = *axiomata*) and principles (*principia*); I'll use 'principles'.

Kabbalistic and Neoplatonic sources

The universality of the categories represented by B–K was a consequence of their origin in the divine attributes or dignities. This is apparent in the First Figure (Fig. 3.2), in which the letter A in the center represents God, from whom emanate the nine divine powers represented by B–K. This is reminiscent of kabbalah, which seems to have influenced Lull; indeed, the *Zohar*, a principal kabbalistic text, was written in Spain while Lull was there (Yates, *AoM*, pp. 178). Later, in the Renaissance, the Lullian letters B–K were explicitly identified with the sefirot (Yates, *AoM*, p. 190), and indeed there is considerable overlap between the Lullian divine attributes (Fig. 3.2 and Table 3.1, "Absolute Principles") and the names of the sefirot (p. 80). Lull's notion of the divine attributes was also influenced by the Neoplatonism of John Scotus Erigena (c.815–77), who identified the divine powers as the first causes of all things; both were influenced by the 5th-century Christian-Neoplatonic *Divine Names* of pseudo-Dionysius (Yates, *AoM*, p. 177). In addition, Lull was probably influenced by a contemporary Spanish kabbalistic practice of meditating on the names of God generated by permutation of the Hebrew alphabet, and by Sufi mystics who also recommended meditation on the names of God (Yates, *AoM*, pp. 178–9). Indeed, Lull thought that these divine attributes were common to Judaism and Islam as well as to Christianity, and so they could be used to convert the former to the latter (Yates, *AoM*, p. 178).

Figures

The terms B–K are the components of the propositions and questions manipulated by the "artist" using Lull's system. The manipulation itself is facilitated by four figures (Figs. 3.2–3.5).

First Figure

The First Figure (Fig. 3.2) shows all the combinations of two terms, where the terms represent the "absolute principles" of Table 3.1. For example, according to context, BF can be interpreted as 'goodness is wise', 'wisdom is good', or even as 'good wisdom' or 'wise goodness'.

Second Figure

The Second Figure (Fig. 3.3) groups the "relative principles" of Table 3.1 into three triads:

difference, concordance, contrariety;
beginning, middle, end;
majority, equality, minority.

These are further divided into three subcategories, represented by the three rings of the figure.

Third Figure

The Third Figure (Fig. 3.4) enumerates the possible combinations of terms. The terms can be interpreted as coming from either the First or Second Figure, so a combination such as BC represents the 12 propositions we get by choosing two of:

good(ness), great(ness), concordance(-ant), difference(-ent)

in either order. Further, two questions are implicit in each proposition, for example:

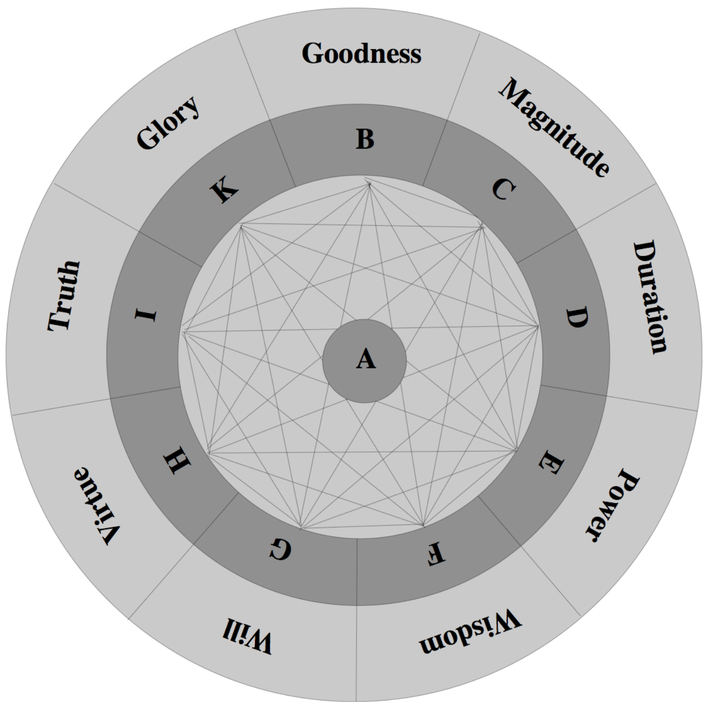


Figure 3.2: Lull's First Figure. The edges connect all possible combinations of any two of the "absolute principles."

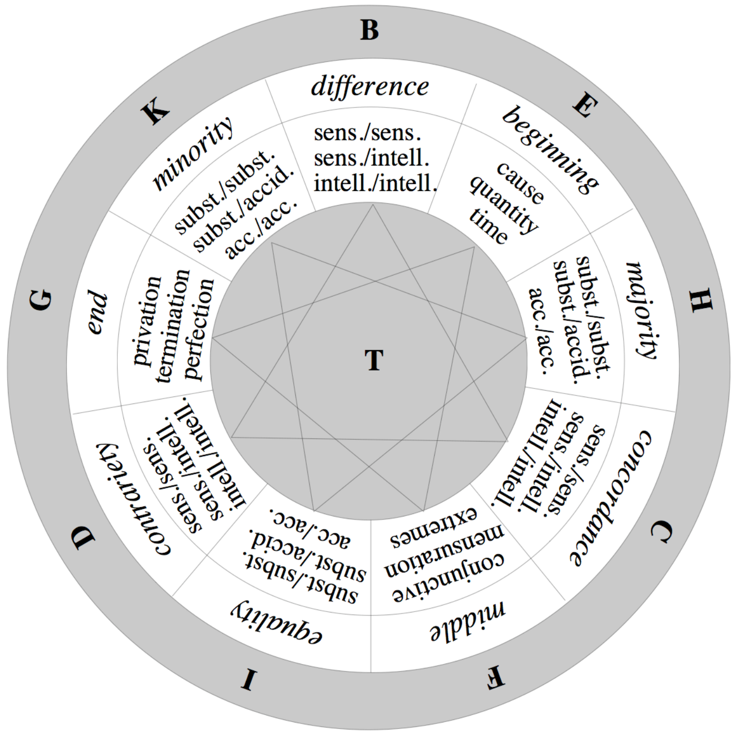


Figure 3.3: Lull's Second Figure. The terms B–K represent the "relative principles." Each of the inner triangles connects three related principles, for example triangle BCD connects difference, concordance, contrariety. Within this are three concentric rings representing different interpretations of the terms. Thus EFG, which in general represents beginning/middle/end, can be interpreted as cause/conjunctive/privation or quantity/mensuration/termination, etc. Abbreviations: acc. = accid. = accident, intell. = intellectual, sens. = sensual, subst. = substance.

BC	CD	DE	EF	FG	GH	HI	IK
BD	CE	DF	EG	FH	GI	HK	
BE	CF	DG	EH	FI	GK		
BF	CG	DH	EI	FK			
BG	CH	DI	EK				
BH	CI	DK					
BI	CK						
BK							

Figure 3.4: Lull's Third Figure. The Third Figure simply represents all combinations of two of the terms B–K.

> Whether greatness is good.
> What is good greatness?

Again, the method is combinatorial.

Fourth Figure

The Fourth Figure (Fig. 3.5) is the only one in the simplified Art that is an actual device, for the two inner rings can be rotated on top of the outer ring. In this way all the possible triples of the terms B–K can be generated. The triples can be used in many ways, for example, to help find the middle term in a syllogism (by trying BCD, BCE, ..., BCK). The Fourth Figure also simplifies the exhaustive investigation of a subject. For example, since B = goodness and C = prudence, we can enumerate all the questions about goodness and prudence (BCB, BCC, BCD, ..., BCK) as follows:

BCB:	Whether prudence is good.	Whether goodness is prudent.
BCC:	What is good prudence?	What is prudent goodness?
⋮	⋮	⋮
BCK:	How is prudence good?	How is goodness prudent?

(Further, Lull specifies two to four *species* of each question, which I'll pass over.)

Is it Logic?

Lull's Art is not a system of logic, nor was it intended as one (as that would "undermine faith"). Although Lull sometimes speaks of the Art *demonstrating* a proposition, he also speaks of it "manifesting" or "revealing" it. Further, the system is not totally mechanical since it requires a trained operator (the "artist") to correctly interpret the configurations of terms. Also, it operates on a background of shared cultural beliefs (especially about God), which are also necessary, for example, in deciding whether a result

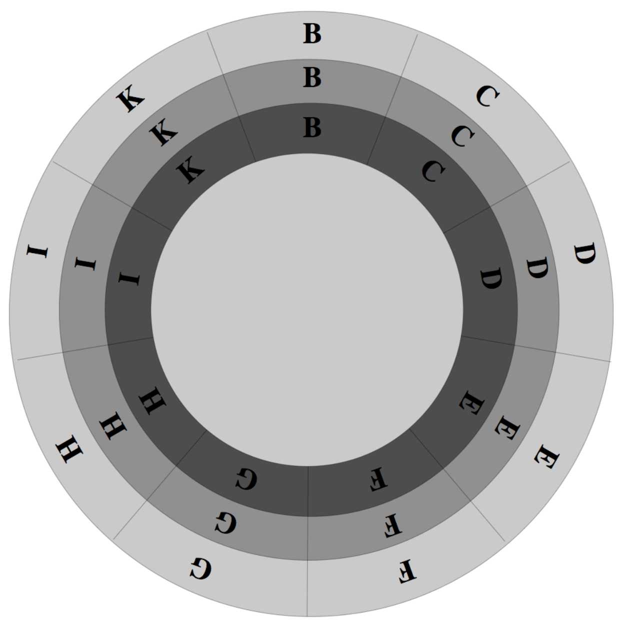


Figure 3.5: Lull's Fourth Figure. This device allows the two inner rings to rotate over the outer ring, thus permitting all the triplets of the terms B–K to be generated. One use of this figure would be to enumerate possible syllogistic arguments.

disagrees with commonly accepted truths. Reaching such an "absurd conclusion" would be the basis for rejecting a hypothesis.

Polemical, not Logical

An interesting question is whether Lull's Art ever produced a result that the artist did not already believe. Indeed, it can be argued that this was not its purpose, for Lull's original goal was to find a means of "converting the infidel" by the word rather than the sword.

Thus the function of the Art was polemical rather that logical. Later, however, Lull seems to have taken the view that the Art could find new truths (see quote on p. 83 regarding the subject of the Art) — but maybe that claim was just another polemical device.

In any case, we must avoid imposing our standards of justification on Lull and his contemporaries. Lull did not *want* to justify Christianity by rational means, since to try to do so was a heresy, "rationalism," because it undermined faith. Indeed, Lullism was twice condemned for this very heresy (in 1376 and 1390), though it was later exonerated (1416). (It is worth reminding ourselves that the papal inquisition began just 30 years before Lull's birth.) In fact, the 1376 condemnation of Lullism was the direct motivation for the publication of the *Directorium inquisitorum*, the definitive manual of inquisitorial methods (Llull, *SWRL*, pp. 71–73).

3.4.3 From Images to Symbols

Decreasing Use of Images

Lullism represents an important stage in the transition from imagistic representations to symbolic representations. As we've seen, a practitioner of the ancient Art of Memory imagined vivid, emotion-laden, sensuous and active images in real or realistic physical locations. By the Middle Ages there was increasing use of abstract geometrical spaces for organizing the images, such as the Zodiac, the Tree of Life and schematic towers, trees and ladders. In many cases iconographic images or even words were used to encode the ideas organized by these structures. (Indeed, realistic images were suspect, because they might lead the pious contemplative away from the path of virtue!)

Increasing Abstraction

Lull replaced images with abstract letters organized into geometrical figures, thus permitting contextual interpretation of the individual symbols and their mechanically generated combinations. (In effect, Lull exchanged the natural motion within the memory images of classical mnemotechnics for mechanical motion between lifeless letters.) In this approach we can see the influence of the kabbalah. The movement from rich imagery to abstract symbols continued in the Renaissance with explicit attempts to combine Lullism and kabbalah and with projects to "mathematize" Lullism by replacing the letters by numbers (Yates, *AoM*, pp. 178–9, 364–5). This trend continued into the early modern period and influenced the birth of science and mirrored the iconoclasm of the Protestant Reformation (Yates, *AoM*, p. 231; see also Sec. 3.4.5 below). It underlies the development of symbolic logic and symbolic AI, although the importance of images and imagistic cognition have been rediscovered in connectionist AI and cognitive science and in studies of imagery in the philosophy of science and engineering.

3.4.4 Significance

Function of Terms and Figures

Lull's Art was inspired and inspirational, and many of its ideas have been adopted by mathematics, symbolic logic, programming languages and other notational systems. By representing the basic concepts by single letters Lull allows propositions to be expressed compactly enough so that their structure is easily seen and so that they can be formally (algebraically) manipulated.[28] In particular, this permits the combinatorial generation of all formulas of a given kind. Further, the terms make explicit the basic concepts of any subject, as well as the relation of composite concepts to the basic ones. We will see that these are basic characteristics of many "ideal" or "philosophical" languages, including the knowledge representation languages used in AI (Sections 4.2, 4.3 and 9.3).

The figures represent potential or actual relationships between the terms. For example, the First Figure shows potential combinations of two terms; the Second Figure shows the triads of relative principles. Other figures, such as the Fourth, actually facilitate mechanical manipulation. Finally, the figures group the basic concepts of the various subjects into knowledge structures, and are mnemonic devices that help the artist to keep the basic concepts and their relationships in mind. (See Section 9.3.)

Mechanization of Discovery

One of Lull's contributions to the theory of knowledge was the idea that the primary truths of a science could be discovered by a two step process: (1) identify the elementary concepts of that science; (2) mechanically generate all possible combinations of the elementary concepts. By actually constructing combinatorial devices he demonstrated his method (and incidentally showed its limitations). Even more important, however, than Lull's specific Art,

[28]Recall (p. 55) Plato's idea that theories are constructed from *elements*, which are analogous to the letters from which syllables are constructed. Lull carries this out quite "literally"!

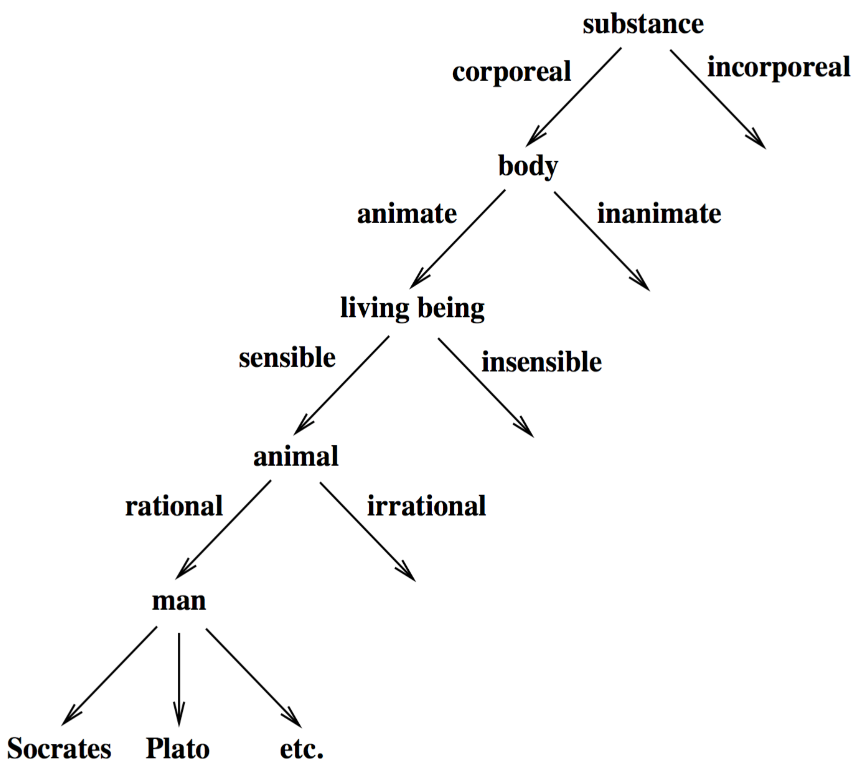


Figure 3.6: The Tree of Porphyry. (This is the common medieval version of the Tree, slightly simplified; his was simpler yet.)

was his vision of a universal method, an inferential process for generating knowledge. This idea found its most direct expression in Leibniz (Section 4.3), but indirectly influenced many others.

3.4.5 Ramus and the Art of Memory

Porphyry's Tree

Before leaving these medieval developments, I'll briefly discuss another one of Porphyry's contributions to philosophy, which was important in the development of logic and symbolic artificial intelligence, the *Tree of Porphyry* (Fig. 3.6). It is a *dichotomy*, that is, a classification in which at each level of the tree we have a binary division, based on contrary predicates, P and non-P. From at least the time of the Sophists a dichotomy was generally considered the most perfect kind of classification.[29] The top of the tree is the *summum genus*, the most general possible class, which in medieval times was taken to be "substance" (*substantia*, that which "subsists through itself"; McKeon, *SMP*, Vol. 2, pp. 499–500). At the bottom of the tree we have individuals, such as Socates and Plato. Therefore, if the entire tree were filled out we would have a complete classification of everything in the universe — animals and people, minerals and vegetables, fire and air, earth and water, angels and devils. Further, it was believed, based on an assertion of Aristotle (*Met. Z* 1032a13–1033a24), that this classification was not arbitrary, but that there was a correct way to do it based on the *true* definition of each class. For example, from the diagram we can see that men (i.e., people) are rational animals and that animals are sensible (e.g., capable of sensation) living beings; therefore men are rational sensible living beings. Continuing this way, we can obtain a compete essential definition of a concept: man is the rational sensible animate corporeal substance. Thus the Tree was taken to be an objective matter of fact representing the actual structure of the universe, "the Great Chain of Being" (Lovejoy, *GCB*; *DHI*, v. 1, pp. 325–35). This is to be expected, for Porphyry and other Neoplatonists, in common with Platonists and Aristotelians, took a *realist* view of universals (recall Sec. 3.2.1).

Class hierarchies

Class hierarchies such as the Porphyry's Tree are still widely assumed in AI and cognitive science, and are used in object-oriented programming languages. They permit concepts to be represented by binary features, corresponding to the yes/no directions one takes on a path from the "root" (*summum genus*) to the concept in question. Unfortunately, although class hierarchies are very tidy, there is psychological evidence that our concepts are not organized this way (discussed in Vol. 2), and they have even been found overly-restrictive for programming (MacLennan, *POPL*, pp. 418–421).

Peter Ramus: 1515–72

The Porphyrean Tree became important in the sixteenth century as a result of the educational reforms of Peter Ramus (Pierre de la Ramée, 1515–72).[30] He was concerned with developing a systematic method for discovering

[29]This is the view taken by the Sophists in their Method of *Diaeresis* or Division (Kerferd, *SM*, pp. 74–75). Later philosophers, including Plato (*Phil.* 16d) and Aristotle (*Part. Anim.* 642b5–644b15), agreed that division must be into the fewest number of subkinds, but thought that this was sometimes more than two. See also Joseph (*IL*, pp. 121–128).

[30]Sources for Ramus include Yates (*AoM*, ch. 10) and Rossi (*LAM*, ch. 5)

and transmitting knowledge (especially in rhetoric, logic, mathematics and geometry), and to this end he combined the arts of memory and logic. Therefore the *Ramean Tree* (or *Ramean Epitome*) proceeds by logical dichotomy from the most general term of any subject matter. In effect the Ramean Tree is an abstract geometrical diagram of the (supposed) essential structure of reality.

Ramism and Method

By organizing knowledge in memory according to this essential structure Ramus hoped to develop a method for preserving, transmitting and discovering knowledge (for classification is often the first step in a scientific investigation). Therefore Ramism contributed to the development of scientific method by Bacon, Descartes and Leibniz.

"Inner Iconoclasm"

With his greater stress on logic, Ramism moved further from the vivid images of the classical art of memory and further in the direction of Lullist abstraction. Yates (*AoM*, pp. 228–9) observes that the Ramist reforms were directed against scholasticism, and therefore were attractive to Protestants. Further, "The extraordinary success of Ramism, in itself a superficial pedagogic method, in Protestant countries like England may perhaps be partly accounted for by the fact that it provided a kind of inner iconoclasm, corresponding to the outer iconoclasm" (Yates, *AoM*, p. 231).

Chapter 4

Thought as Computation

4.1 Hobbes: Reasoning as Computation

The Mechanical Philosophy

In the seventeenth century Pierre Gassendi (1596–1655), René Descartes (1596–1650) and others developed a *mechanical philosophy* of nature, which explained physical processes in terms of mathematically describable material properties: size, shape, mass, etc. Everything is nature was supposed to be reducible to these terms, but Descartes was careful to draw a sharp line between mind and matter, subjecting the material world to mechanistic analysis while leaving mind — specifically the rational soul — to religion and the theologians. In this way he avoided the anti-religious tendencies of the mechanical philosophy, but others, such as Thomas Hobbes, were not so shy and followed the mechanistic philosophy to its materialistic and atheistic conclusions.

Thomas Hobbes: 1588–1679

Perception

In particular, Hobbes took a very mechanistic view of cognition.[1] The causes of sensations are external objects, which either directly or indirectly "press" on our sense organs. The nerves in turn pass this pressure on to the brain, where there is a counterpressure, which constitutes sensation. Sensible qualities are nothing but *motions*, whether in the body that produces them, or in our nervous systems.

Illustrative Quotation

Thus:

> All which qualities called *Sensible*, are in the object that causeth them, but so many several motions of the matter, by which it presseth our organs diversely. Neither in us that are pressed, are

[1] The source for this section is Hobbes (*Lev.*, Part 1, Chs. 1–5).

> they anything else, but diverse motions; (for motion, produceth nothing but motion.) (Hobbes, *Lev.*, Part 1, Ch. 1, p. 3)

Mental Discourse

According to Hobbes thought takes place when ideas follow upon one another in a *train*. This succession may be undirected, as when our thoughts wander, or they may be directed. Hobbes claims that directed trains of thought are of two kinds:

1. When the effect is known and we seek the causes; all animals exhibit this kind of thought in planning their actions.

2. When the cause is known and we seek the effects that can arise from it; this kind of thought is peculiar to man.

Hobbes says that trains of thought constitute a *mental discourse* analogous to the *verbal discourse* that occurs when we speak.

Backward Chaining vs. Forward Chaining

Hobbes' two kinds of trains of thought correspond to the two inference strategies used in most modern automated reasoning systems. *Backward chaining* is reasoning backward from the desired conclusion to find premisses that will imply it. *Forward chaining* is reasoning forward from the premisses to see the conclusions to which they lead.

Language of Thought

The idea that thinking is essentially talking to oneself, and that therefore there is a *language of thought*, has a long history in philosophy and psychology. The idea is still vigorously defended, for example, by Jerry Fodor (*LT*, *Rep.*, *PS*; see also Sec. 9.3.6).

Nature of Language

Speech Reflects Mental Discourse

Hobbes claims that the purpose of speech is to transfer mental discourse to verbal discourse, that is, to turn a train of thoughts into a train of words (spoken or written). Two major reasons for doing this are:

1. To record our thoughts in a stable form, so that we can be reminded of them.

2. To convey our thoughts to others.

Indeed, Hobbes thinks that without language it's impossible to know general truths at all. For example, without words such as *triangle* and *two* we could not know that the angles of a triangle sum to two right angles.

Necessity of Definitions

The function of language is defeated if our words do not accurately reflect our thoughts. Therefore, Hobbes argues that anyone claiming true knowledge must begin by setting down definitions, just as is done in geometry. The definitions give the terms definite and fixed meanings. Just as accountants, in balancing their books, cannot expect to get correct results unless their starting figures are accurate, so philosophers cannot expect to obtain correct results unless their definitions are accurate.

Illustrative Quotation

Thus:

> Seeing that *truth* consisteth in the right ordering of names in our affirmations, a man that seeketh precise *truth*, had need to remember what every name he uses stands for; and to place it accordingly; or else he will find himselfe entangled in words, as a bird in lime-twiggs; the more he struggles, the more belimed. (Hobbes, *Lev.*, Pt. 1, Ch. 4, p. 15)

Nominalism

Consistent with his materialistic views, Hobbes denied the reality of ideas (in a Platonic sense), and took a nominalist view of concepts (recall Sec. 3.2.1). Thus, while we must be careful to define our terms, the definitions are ultimately arbitrary and a matter of convenience, rather than reflecting an underlying reality.

Nature of Reason

Reasoning as Calculation

For Hobbes, reasoning is a process of calculation in which (properly defined) words are manipulated as tokens. He repeatedly draws the analogy with accounting: just as we manipulate numbers or tokens to balance our accounts, so we manipulate words to reason. But this is more than just an analogy, for he says,

> When a man *Reasoneth*, hee does nothing else but conceive a summe totall, from *Addition* of parcels; or conceive a Remainder, from *Subtraction* of one summe from another: which (if it be done by Words,) is conceiving of the consequence of the names of all the parts, to the name of the whole; or from the names of the whole and one part, to the name of the other part. (Hobbes, *Lev.*, Pt. 1, Ch. 5, p. 18)

Hobbes has been criticized for taking the computational view too literally, but he deserves more credit, for it's clear that he has more than numerical sums and differences in mind. What he intends are *formal* sums and differences, that is, the synthesis and analysis of symbol structures. He makes this clear by observing, for example, that geometrical figures are "sums" of lines, angles, etc. More to the point, he notes that propositions are sums of terms, that syllogisms are sums of propositions, and that demonstrations are sums of syllogisms. As he says, addition and subtraction "are not incident to Numbers onely, but to all manner of things that can be added together, and taken one out of another." (Hobbes, *Lev.*, Pt. 1, Ch. 5, p. 18) Thus, for Hobbes, reasoning is the manipulation of structures of symbols (words) by formal analytic and synthetic processes.

Illustrative Quotations

For example, he says,

> For REASON, in this sense, is nothing but *Reckoning* (that is, Adding and Subtracting) of the Consequences of generall names agreed upon, for the *marking* and *signifying* of our thoughts... (Hobbes, *Lev.*, Pt. 1, Ch. 4, p. 18)

And also:

> By *ratiocination* I mean *computation*. Now to compute, is either to collect the sum of many things that are added together, or to know when one thing is taken out of another. (Hobbes, *Elem. Phil.*, Sect. 1, de Corpore 1, 1, 2)

On the use of words as tokens, Hobbes says,

> For words are wise mens counters, they do but reckon by them: but they are the mony of fooles, that value them by the authority of an *Aristotle*, a *Cicero*, or a *Thomas*, or any other Doctor whatsoever, if but a man. (Hobbes, *Lev.*, Pt. 1, Ch. 4, p. 15)

Science

Nevertheless he follows Aristotle (Section 2.5.2) in arguing that science is the result of accurate definition of names together with the formal connection of assertions in a rigorous deductive structure. This is a result of industry rather than experience.

> To conclude, The Light of humane minds is Perspicuous Words, but by exact definitions first snuffed, and purged from ambiguity; *Reason* is the *pace*; Encrease of *Science*, the *way*; and the Benefit of man-kind, the *end*. (Hobbes, *Lev.*, Pt. 1, Ch. 5, pp. 21–22)

Experience vs. Science

Although Hobbes is obviously in favor of a computational approach to reasoning, he acknowledges an important distinction:

> [T]he Latines did always distinguish between *Prudentia* and *Sapientia*; ascribing the former to Experience, the later to Science. (Hobbes, *Lev.*, Pt. 1, Ch. 5, p. 22)

Like Socrates, Hobbes says that the advantage of sapience (scientific knowledge) is that it is infallible. In contrast to Socrates (Section 2.4.3), he recognizes the important pragmatic value of the prudence (practical wisdom) that may come with experience, especially when it is combined with sapience. Further he recognizes the danger of sapience without prudence:

> But yet they that have no *Science*, are in better, and nobler condition with their naturall Prudence; than men, that by misreasoning, or by trusting them that reason wrong, fall upon false and absurd generall rules. (Hobbes, *Lev.*, Pt. 1, Ch. 5, p. 21)

Therefore, to avoid rationalism, reason needs to rest on a firm foundation of experience.

Summary

We have seen that for Hobbes cognition is nothing more than a kind of matter in motion. We have also seen that he takes reasoning to be a kind of calculation. An implication of these two claims, which Hobbes apparently doesn't see but which others will, is that it ought to be possible to build a machine that reasons by calculation. That is, since cognition is but matter in motion, no special nonmaterial substance (e.g., soul) is prerequisite to reasoning. Nevertheless, it will be about 200 years before a reasoning machine is actually constructed (Section 4.5).

4.2 Wilkins: Ideal Languages

John Wilkins: 1614–1672

It seems that for as long as people have written about language, they have complained of the problems that arise from the diversity of natural languages. Later, as they studied the ways language is used in argument (rhetoric and logic), they became aware of the imperfections of natural languages, especially their ambiguity and lack of logical structure, and we have seen that the schoolmen tried to refine scholastic Latin into a language of logic. In the century preceding the Age of Reason, this discontent precipitated a number of projects to design *ideal languages*. The philosophers who worked on this problem include Bacon, Descartes, Mersenne and Leibniz. Indeed, since that time approximately 500 ideal languages have been defined, and the activity continues to this day. Here we will discuss briefly the "Real Character" of John Wilkins, one of the most fully developed systems, and one which influenced Leibniz in his development of logical calculi (Section 4.3).[2]

Assumptions

Several assumptions, typical of the seventeenth century, underlie Wilkins' effort and most of the others. One was the *realist* assumption that the basic concepts are the same for all people regardless of the language they speak.[3] Thus, an English speaker might use a phrase where a German speaker uses a single word, or vice versa, but they both denote the same concept. Certainly, some languages have concepts that are unknown to speakers of some other languages (e.g., the name for an animal unknown to the latter), but it was assumed that this concept could be easily grasped if the need arose. Experimental evidence supports the idea that languages may facilitate or impede making certain distinctions, but that they do not impose their own inescapable "reality" (Crystal, *CEL*, pp. 14–15).[4]

Second, they assumed that underlying all the peculiarities of the grammars of individual languages there are certain universal principles — a universal grammar — that reflect the laws of thought:

> As men do generally agree in the same Principle of Reason so do they likewise agree in the same *Internal Notion* or *Apprehension of things*. (Wilkins, *RC*, §I.5.2, p. 20)

An example of such a principle might be the subject-predicate relation analyzed by Aristotle. In fact, there seem to be no nontrivial characteristics that *all* languages have in common, and few that even *most* languages have in common. One near universal is that grammatical subjects precede grammatical objects in over 99% of the languages investigated. On the other hand, even so simple a claim as "all languages have words" is problematic (Crystal, *CEL*, pp. 84–85). It is apparent that neither of these assumptions can be taken for granted.

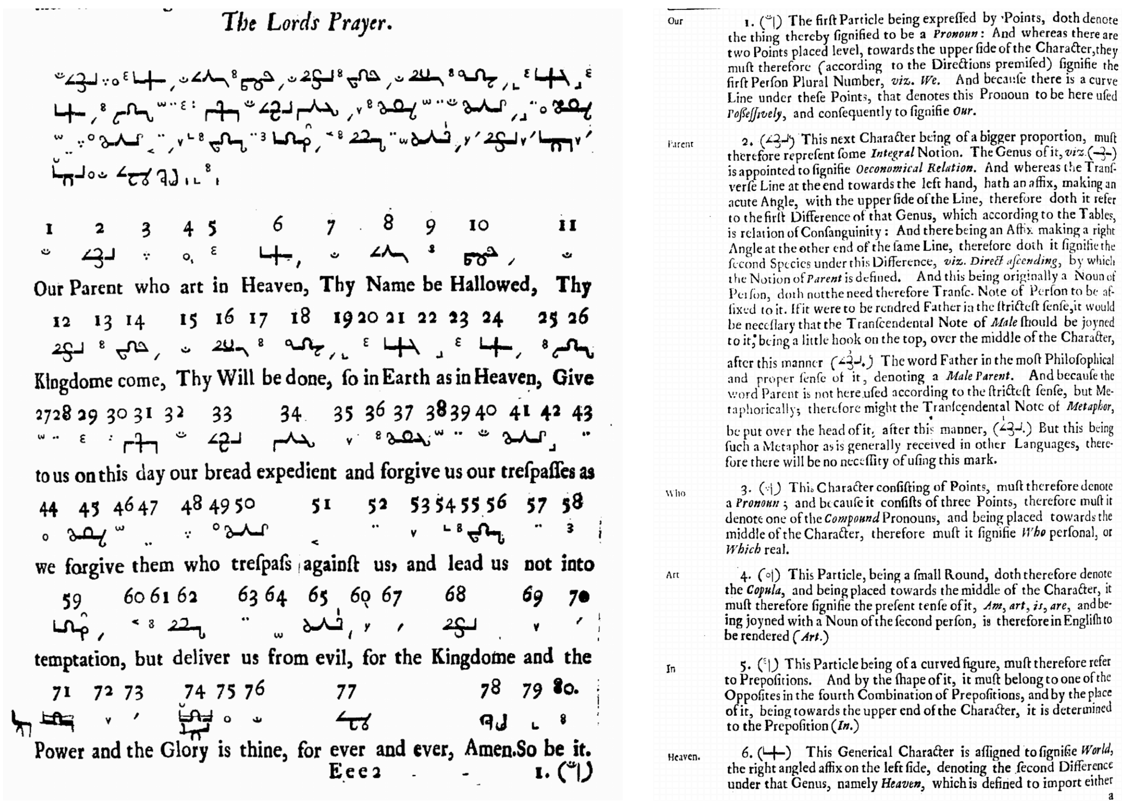
The Real Character

One of the defects Wilkins perceived in natural languages was that there is no systematic relation between the forms of words and their meanings. In addition to some words having several meanings, and there being several words for the same concept, there is no way to see the logical relation between words from their written or spoken forms. For example, since men, dogs and fish are all animals, it would seem reasonable that some part of the words for

[2]Principle secondary sources for this section are: Vickers (*ES*, Ch. 9), and Ellegård (1973, pp. 668–669). See also Eco (*SPL*), Rossi (*LAM*) and Large (*ALM*).

[3]Recall the *Port Royal Grammar* and *Logic*, p. 70.

[4]One popular book (Rheingold, *THWFI*) collects "untranslatable words and phrases" from more than forty languages.


The Lords Prayer.

1 2 3 4 5 6 7 8 9 10 11
Our Parent who art in Heaven, Thy Name be Hallowed, Thy
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Kingdome come, Thy Will be done, ſo in Earth as in Heaven, Give
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
to us on this day our bread expedient and forgive us our treſpaſſes as
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
we forgive them who treſpaſs againſt us, and lead us not into
59 60 61 62 63 64 65 66 67 68 69 70
temptation, but deliver us from evil, for the Kingdome and the
71 72 73 74 75 76 77 78 79 80.
Power and the Glory is thine, for ever and ever, Amen. So be it.
Eee2 1.

Our 1. The firſt Particle being expreſſed by Points, doth denote the thing thereby ſignified to be a *Pronoun*: And whereas there are two Points placed level, towards the upper ſide of the Character, they muſt therefore (according to the Directions premiſed) ſignifie the firſt Perſon Plural Number, *viz. We.* And becauſe there is a curve Line under theſe Points, that denotes this Pronoun to be here uſed *Poſſeſſively*, and conſequently to ſignifie *Our*.

Parent 2. This next Character being of a bigger proportion, muſt therefore repreſent ſome *Integral* Notion. The Genus of it, *viz.* is appointed to ſignifie *Oeconomical Relation.* And whereas the Tranſverſe Line at the end towards the left hand, hath an affix, making an acute Angle, with the upper ſide of the Line, therefore doth it refer to the firſt Difference of that Genus, which according to the Tables, is relation of Conſanguinity: And there being an Affix making a right Angle at the other end of the ſame Line, therefore doth it ſignifie the ſecond Species under this Difference, *viz. Direct aſcending*, by which the Notion of *Parent* is defined. And this being originally a Noun of Perſon, doth not the need therefore Tranſc. Note of Perſon to be affixed to it. If it were to be rendred Father in the ſtricteſt ſenſe, it would be neceſſary that the Tranſcendental Note of *Male* ſhould be joyned to it, being a little hook on the top, over the middle of the Character, after this manner The word Father in the moſt Philoſophical and proper ſenſe of it, denoting a *Male Parent.* And becauſe the word Parent is not here uſed according to the ſtricteſt ſenſe, but Metaphorically; therefore might the Tranſcendental Note of *Metaphor*, be put over the head of it, after this manner, But this being ſuch a Metaphor as is generally received in other Languages, therefore there will be no neceſſity of uſing this mark.

Who 3. This Character conſiſting of Points, muſt therefore denote a *Pronoun*; and becauſe it conſiſts of three Points, therefore muſt it denote one of the *Compound* Pronouns, and being placed towards the middle of the Character, therefore muſt it ſignifie *Who* perſonal, or *Which* real.

Art 4. This Particle, being a ſmall Round, doth therefore denote the *Copula*, and being placed towards the middle of the Character, it muſt therefore ſignifie the preſent tenſe of it, *Am, art, is, are*, and being joyned with a Noun of the ſecond perſon, is therefore in Engliſh to be rendered (*Art.*)

In 5. This Particle being of a curved figure, muſt therefore refer to Prepoſitions. And by the ſhape of it, it muſt belong to one of the Oppoſites in the fourth Combination of Prepoſitions, and by the place of it, being towards the upper end of the Character, it is determined to the Prepoſition (*In.*)

Heaven. 6. This Generical Character is aſſigned to ſignifie *World*, the right angled affix on the left ſide, denoting the ſecond Difference under that Genus, namely *Heaven*, which is defined to import either
a

Figure 4.1: Example of Wilkin's Real Character (Wilkins, *RC*, pp. 395–6). Additional examples can be found in Vickers (*ES*, pp. 194–195).

these concepts would mean *animal*, but this is not the case. Thus the first part of Wilkins' project was to define a "real character," or philosophical notation whereby the symbols for concepts reflected their logical relations (which are objective, according to *realism*). In this he was inspired by Chinese characters, but he intended his symbols to be more rational than the Chinese, which tend to be constructed metaphorically. The spoken version of Wilkins' language was to be based on these written characters (Fig. 4.1).

Taxonomy of Concepts

The first step in designing the Real Character was to develop a taxonomy of basic concepts. (In this he was aided by the "analytic dictionaries" that were becoming popular at that time.) Wilkins first divided knowledge into 40 domains. Within each of these domains were about six genera, and within each genus about ten species. Thus he had a taxonomy of about 2400 basic concepts, his "principal words." If you want an idea of what such a taxonomy is like, look in an older (i.e., nonalphabetized) edition of *Roget's Thesaurus* and you will see a system much like Wilkins' (and in fact inspired by it). Of course, 2400 concepts are not an adequate vocabulary, so Wilkins provided two mechanisms for building additional words. One was "compounding," the combination of existing words to achieve a compound meaning. The other was a system of "particles" that were attached to the main symbol like Arabic or Hebrew vowel signs (Masoretic points). The particles modify the meanings of symbols in systematic ways.

Example Particle

One of Wilkins' particles, for example, enlarges the sense of a word, so that the new concept is related to the original concept metaphorically. Here are some of Wilkins' examples of root concepts and their metaphorical modifications:

Light	→	Evident, plain
Dark	→	Mystical, obscure
Ripe	→	Perfect
Shining	→	Illustrious
Raise	→	Prefer, advance

Other examples are in Vickers (*ES*, p. 192).

Metaphors

It is of course an interesting question to what extent these metaphors transcend cultures. Cherry (*OHC*, p. 74) says that the use of washing and cleanliness metaphors to refer to absence of sin is peculiar to cultures with a strong Christian influence, and that the notion of "the mind's eye" is not familiar to the Chinese. In the ancient Greek and Roman worlds it appears (Onians, *OET*, pp. 30–37) that diminished consciousness, as in sleep and intoxication, was thought of as *wet*, and that clear consciousness was *dry*. For example, a common Homeric word for 'prudent' (*peukalimos*) seems to refer to the dryness of the bronchial tubes, since they were thought to be the seat of consciousness and intelligence. Also, this is why *Lêthê* (Forgetfulness) is a liquid and is drunk, and why a forgetful person might be accused of having a "wet" memory. On the other hand, many metaphors may have a biological, and hence transcultural, basis (Johnson, *BiM*; Lakoff, *WFDT*; Lakoff & Johnson, *MWLB*).

Universal Grammar

Wilkins distinguished the "instituted" or "particular" grammars of natural languages from the "natural" or "universal" grammar that he believed to underlie them all:

> *Natural* Grammar (which may likewise be stiled Philosophical, Rational, and Universal) should contain all such Grounds and

> Rules as do naturally and necessarily belong to the Philosophy of letters and speech in the *General*. (Wilkins, *RC*, §III.1.1, p. 297)

Universal grammar is perhaps tied to the universal laws of thought, and perhaps even reflects a language of thought. Whether such a things exists is still debated (Section 9.3.6).

Wilkins' intention was that his language have no grammatical rules beyond those of natural grammar. Thus his language would not require the learning of additional grammar, since everyone already knows the universal grammar! For this reason he believed that to learn his language it was only necessary to learn the "principal words" and particles, and on this basis he estimated that his language was 40 times easier to learn than Latin (Vickers, *ES*, p. 197).

Summary

Wilkins' Real Character was a failure; indeed, it was something of an embarrassment to the Royal Society, of which he was the first (acting) president. Nevertheless, the prospect of an ideal language continues to attract. On one hand this has led to the development of various "international" languages, such as Esperanto. These are intended to be complete languages suitable for all the purposes for which natural languages are used. On the other hand it has led to the development of languages and notations for special applications, such as symbolic logic. These languages are not intended to be complete; they are generally adequate only for the expression of propositions (declarative statements), and often only within a restricted domain. We will explore this line of development next, when we investigate Leibniz' contributions.

Loglan: a contemporary ideal language

I will briefly describe *Loglan*, a contemporary ideal language. (Readers uninterested in Loglan should skip to p. 114.) I have chosen Loglan because (1) it is much less familiar than languages such as Esperanto, and (2) it has many interesting characteristics relevant to this book. James Cooke Brown began the design of Loglan in December 1955 in order "to test the Sapir-Whorf hypothesis that the structure of language determines the boundaries of human thought" (Brown, *L1*, p. 1). 'Loglan' stands for 'logical language', and the simplest description of Loglan is *speakable predicate logic*. However, it goes beyond logic by providing metaphorical and other nonlogical means of expression.

Loglan has a number of attractive features. First, it is a small language, having 257 simple grammar rules, 834 basic predicates (its root vocabulary), and 120 function words. Second, it has many formal transformation rules, which, among other things, facilitate formal logical derivation. Third, it strives for cultural neutrality by drawing its phonetic patterns and word roots from the eight most widely spoken languages at that time (English, Mandarin Chinese, Hindi, Russian, Spanish, French, German, Japanese). Together they account for approximately 80% of the world's population. For example, the Loglan word for 'blue' is 'blanu', which will seem familiar to speakers of four of these languages:

$$\text{blanu} \iff \begin{array}{ll} \text{blue} & \text{(English)} \\ \text{lan} & \text{(Chinese)} \\ \text{bleu} & \text{(French)} \\ \text{blau} & \text{(German)} \end{array}$$

It also has affinities to Hindi 'nila', Spanish 'azul' and Russian 'galuboi'. Fourth, Loglan also strives for cultural neutrality by "metaphysical parsimony," that is, by building into the language few assumptions about the world.[5] For example, there is no obligatory tense system (such as English has), and no obligatory epistemic inflection (such as Hopi has), but either or both may be used if desired. Fifth, Loglan is more expressive than English or other natural languages (see below for examples). Finally, Loglan boasts almost complete freedom from ambiguity, although Brown admits that this might make it unsuitable for poetry.

Some other noteworthy features of Loglan are: (1) spoken punctuation (including parenthesis and quotation marks); (2) no distinction between nouns, verbs, adjectives and adverbs; they are all predicates; (3) word boundaries can be unambiguously determined; a word's part of speech (predicate, conjunction, etc.) can be determined by its phonetic pattern; (4) explicit scope for conjunctions and quantifiers; (5) free variables and various kinds of quantified variables for individuals.[6]

I'll present a few examples to illustrate some features of the language. First, a fairly complex sentence (Brown, *L1*, p. 226):

[5] However, the objection can be made that structuring the language around predicate logic is in itself a substantial metaphysical commitment!

[6] Predicate variables have also been proposed (MacLennan, PV).

> Mi pa ferlu Inukou ki la Djan pa kanvi mi jia kamla kimoi da pa setfa le banla ta
>
> I fell because John saw me coming (and) therefore he (was motivated to) put the banana there.

This should be easy to unravel with the following vocabulary:

mi = I	kamla = come
pa = past	kimoi = therefore (motivational)
ferlu = fall	da = X/he/she/it
Inukou = physical causation	setfa = set
la Djan = the one named 'John'	le = a thing of type
kanvi = see	banla = banana
jia = who/which/that	ta = that/those/there

The word 'ki' has no direct translation; it is a kind of bracket, which defines the scope of the following 'kimoi' (i.e., it delimits the motivation).

As an example of the ambiguity of English, Brown (*L1*, App. A) lists *seventeen* meanings of "pretty little girls school"; try saying it with different inflections. All seventeen meanings can be expressed exactly in Loglan; I show just four, which use 'ge' to group the following modifiers (bilti = beautiful; cmalo (*shmalo*) = small; nirli = girl; ckela (*shkela*) = school):

Da bilti cmalo nirli ckela	X is a beautifully small girls' school.
Da bilti ge cmalo nirli ckela	X is beautiful for a *small* girls' school.
Da bilti cmalo ge nirli ckela	X is beautifully small for a girls' school.
Da bilti ge cmalo ge nirli ckela	X is beautiful for a small (type of) girls' school.

You should be able to follow the remaining examples (all from Brown, *L1*) with minimal explicit vocabulary.

The following example shows how Loglan distinguishes personal and material supposition:

La Djan corta purda	John is a short word.	**false!**
Li Djan corta purda	'John' is a short word.	**true!**

Some ways of avoiding numerical ambiguity:

Kambe leva fefe galno veslo	Bring those *fifty-five* ... gallon-cans.
Kambe leva feni fera galno veslo	Bring those fifty ... five-gallon-cans.
Kambe leva ri fefera galno veslo	Bring those fifty-*five*-gallon cans.

The following show ways of talking about properties:

Mi clivu lo gudbi	I love good things.
Mi clivu lo pu gudbi	I love the property that good things have.
Mi clivu lo po gudbi	I love good states-of-affairs.
Mi clivu lo zo gudbi	I love all the quantities of goodness in good things.

The following illustrate indefinite and definite quantifiers (to = 2, te = 3, si = at most, ve = 9, ba = some x, be = some y):

levi to fumna ga corta leva te mrenu	These two women are (all) shorter than (each of) these three men.
Sive le botci pa kamla le sitci	At most nine of the boys came from the city.
Ba no be: ba corta be	There is an x such that there is no y such that x is shorter than y.

For control of the scope of quantifiers, compare:

Re le mrenu: da merji anoi farfu	For most of the men, X is married if a father.
Re le mrenu ga merji anoi farfu	Most of the men are married if most of the men are fathers.

These show control of the scope of logical connectives:

Da pa rodja Inunokou tu no pa cutri durzo da Inukou tu no pa danza le po da rodja	It grew although you didn't water it, *because* you didn't want it to grow.
Da pa rodja Inunokou ki tu no pa cutri durzo da kinukou tu no pa danza le po da rodja	It grew, *although* you didn't water it because you didn't want it to grow.

There are many logical operators; for example 'cu' denotes logical *independence*:

Da forli cu kukra prano	X is a strong — whether fast or not — runner.

Needless to say, these are just a few, isolated samples of Loglan; see Brown (*L1*, *L4&5*) for more information and Brown (Loglan) for a brief introduction. The Loglan Institute maintains an Internet site with extensive reference material.

4.3 Leibniz: Calculi and Knowledge Representation

> The General Science is nothing else than a science of cognition, ... not only a logic, but an art of discovery, a Method or manner of ordering, a Synthesis and Analysis, a Pedagogy or science of teaching, ... an Art of Memory and Mnemonics, an *Ars Characteristica* or *Symbolic Art*, an *Ars Combinatoria*, ... a philosophical Grammar, a Lullian Art, a Cabbala of the Wise, a Natural Magic. ... In short all sciences will be here contained as in an Ocean.[7]
>
> — Leibniz, *Introductio ad encyclopædiam arcanam* (Couturat, *Opl. Leib.*, pp. 511–12; cf. Yates, *AoM*, p. 370; Rossi, *LAM*, p. 191)

> Anyone who knows me only by my publications does not know me at all.
>
> — Leibniz (Coudert, *L&K*, p. 2)

4.3.1 Chinese and Hebrew Characters

Leibniz: 1646–1716

Gottfried Wilhelm von Leibniz has a long list of intellectual accomplishments. In addition to a career as a jurist, he invented the differential and integral calculi (at about the same time as Newton), made important contributions to all fields of philosophy, and even constructed an early calculating machine. Here we will be mainly concerned his investigations into knowledge

[7] *Scientia Generalis nihil aliud est quam Scientia [cogitandi] ..., quæ non tantum <complectitur> Logicam hactenus receptam, sed et artem inveniendi, et Methodum seu modum disponendi, et Synthesin atque Analysin, et Didacticam, seu scientiam docendi; Gnostologiam, quam vocant, Noologiam, Artem reminiscendi seu Mnemonicam, Artem characteristicam seu symbolicam, Artem Combinatoriam,* Artem Argutiarum, *Grammaticam philosoφicam: Artem Lullianam, Cabbalam sapientum, Magiam naturalem. ... Non multum interest quomodo Scientas partiaris, sunt enim corpus continuum quemadmodum Oceanus. (Couturat, Opl. Leib., pp. 511–12)*

representation and inferential calculi.[8] Before considering these topics, however, I will mention two of Leibniz's interests that influenced his ideas about thought and language.

Binary Numbers and the I Ching

Like Wilkins, Leibniz was impressed by Chinese ideographic writing, which he saw as a more direct representation of thought than European phonetic scripts.[9] The Jesuits had begun missionary activity in China in the early sixteenth century, which led to the appearance of books about China and translations of Chinese works into Latin. Leibniz had read books about China and Chinese writing by the enthusiastic but unscholarly Jesuit Athanasius Kircher (1602–80), which, among other things, attempted to connect Chinese characters to Egyptian hieroglyphics (which had not been deciphered at that time). (Kircher was an important contributer to the Hermetic "magical philosophy," which was popular at that time, and which contributed to the birth of modern science; it will be discussed in Sec. 5.2). Another Jesuit, Joachim Bouvet (c.1656–1730), showed Leibniz a translation of an important Chinese philosophical text, the *I Ching* (*Book of Changes*), which also made a deep impression. Leibniz had already invented the binary number system, which represents all numbers by means of just two symbols, '0' and '1' (corresponding to false and true), and so he was astonished to discover the same pattern in the *I Ching*, which is based on 64 *hexagrams*, each comprising six lines of two possible types (broken *yin* or solid *yang*). These opposites correspond to all the polarities (true / false, male / female, light / dark, etc.), and so the *I Ching* seemed to be an analysis of reality into its elementary logical constituents. (Recall also the Pythagorean Table of Opposites, p. 28.)

Van Helmont and Kabbalah

Kircher and Bouvet, along with other Hermetic philosophers (Sec. 5.2), believed that Chinese philosophy represented the same "ancient theology" (*prisca theologia*) that had been taught to Moses by God, but that this original philosophy had become confused over the centuries. Furthermore kabbalah seemed like a promising approach to recovering this ancient philosophy, and so it is unsurprising that Leibniz was interested in it.

Although Leibniz was aware of gematria and other kabbalistic practices, his interest became more acute in 1667 when he began reading the *Short Sketch of the Truly Natural Alphabet of the Holy Language* by Franciscus

[8] One of the best collections of Leibniz's logical papers is Parkinson (*LLP*). Description of his logical calculi can also be found in Kneale & Kneale (*DL*) and in Styazhkin (*HML*). Leibniz's interpretation of his calculi is discussed in Rescher (LILC).

[9] The principal source for this information is Dusek (*HIP*, ch. 11).

Mercurius van Helmont (1614–99), son of the famous chemist (and alchemist) Jan Baptist van Helmont (1577–1644).[10] This book argued that in the original Adamic language, the shapes and sounds of the letters reflected the real nature of things, but that this accuracy had been lost in Biblical Hebrew, which dated from after "The Fall."[11] Leibniz and van Helmont met in 1671 and discussed alchemy, for they both were practicing alchemists (as were their contemporaries Newton, Boyle and Locke; see Sec. 5.2.3).[12] Van Helmont introduced Leibniz to his friend Christian Knorr von Rosenroth (1636–89), whom he was assisting in the preparation of *Kabbala Denudata* (*Kabbalah Unveiled*), the first comprehensive Latin translation of kabbalistic texts. Leibniz and van Helmont became friends and continued to discuss kabbalah until the latter's death in 1691.[13] They saw language, and in particular the words and letters of the original Hebrew of Adam, as the link between mind and matter. Therefore they rejected Cartesian dualism, which separated mind from matter, and they rejected the nominalism of Hobbes, Locke and others, because reality was encoded in the "natural alphabet of the holy language." These views reflect Hermetic and Neoplatonic philosophy, which they shared with many of their contemporaries, according to which archetypal ideas emanated from God in a "great chain of being" and gave form to the natural world and everything in it. As a consequence there was a harmony of analogous or proportionate structures on all the levels of being, including that of the "natural language" (or "language of nature"), in which the Book of Nature was written. In particular, Leibniz believed that there was a close connection between reasoning and the characters in which that reasoning is represented, which brings us back to knowledge representation and inference.

The effects of Leibniz' interests in Chinese philosophy and kabbalah can be discerned in his *monadology*, the philosophical system for which he is best known (but not our concern here).

[10] The principal source for this topic is Coudert (*L&K*).

[11] Astonishingly, Leibniz eventually came to believe that the language of Adam was closer to *German*!

[12] Leibniz was still talking about alchemy on his deathbed, to the dismay of the attending priest, who thought that he should be more concerned about the state of his soul!

[13] Indeed, Leibniz "ghost-wrote" van Helmont's last book, *Some Premeditate and Considerate Thoughts upon the Four First Chapters of the First Book of Mosis [sic] called Genesis*, which deals with kabbalah.

4.3.2 Knowledge Representation

Analysis of Concepts into Elements

Leibniz claims that "all things, which exist or can be thought of are in the main composed of parts, either real or at any rate conceptual" (*A.6.1*, 177; *LLP* 3).[14] Therefore all thought involves the analysis of concepts into their parts and the synthesis of these parts into new combinations. As we saw, both Lull and Hobbes had already expressed this view (Sections 3.4 and 4.1). Leibniz was inspired by Lull's ambitious system, but thought it unworkable. Hobbes had also caught his attention: "Thomas Hobbes, everywhere a profound examiner of principles, rightly stated that everything done by our mind is a *computation*" (*A.6.1*, 194; *LLP* 3). Like them, Leibniz believed that the analysis of concepts had to come to an end at some point, when certain elementary or *atomic* concepts were reached. Citing Aristotle, he says "these final terms are understood, not by further definition, but by analogy" (*A.6.1*, 195; *LLP* 4).[15]

Alphabet of Thought

Leibniz thought that the analysis of concepts into their elements could be used as the basis for a universal writing and an *alphabet of thought*. We have already seen (Section 4.2) how Wilkins designed a language whose notation embodied a scientific taxonomy of concepts. Leibniz approved of Wilkins' project, but thought he could do better by designing a notation that embodied the very logical structure of concepts. "If [the characters] are correctly and ingeniously established, this universal writing will be as easy as it is common, and will be capable of being read without any dictionary; at the same time, a fundamental knowledge of all things will be obtained." (*A.6.1*, 202; *LLP* 11) These were ambitious plans. Although Leibniz never completed the design of this language, we'll see that some of the knowledge representation and processing techniques he developed anticipated those currently in use in AI.

Prime Decomposition of Concepts

Leibniz believed that every concept could be analyzed into a number of atomic concepts, and conversely that every concept was determined by the

Prime and Composite Concepts

[14]References beginning *A* are to series, volume and page number of the Academy edition (Leibniz, Academy) of Leibniz's works, and references beginning *C* are to Leibniz (Couturat). All selections quoted are also available in Parkinson (*LLP*), and will also be cited by their page number therein (marked *LLP*).

[15]The phrase Aristotle uses (*Metaphysics* 1048a36–38), τὸ ἀνάλογον (*to analogon*), means analogy or proportion. Since grasping the analogy cannot depend on a further conceptual analysis, perhaps Aristotle means by this an immediate grasp of the similarity between two situations; he does say that the analogy "can be abstracted from the particular cases."

atomic concepts that composed it. Leibniz was also an excellent mathematician, and quickly recognized the similarity between prime numbers and atomic concepts. Ordinary concepts are like composite numbers: they can be divided in just one way into smaller components. But analysis must stop when the indivisible elements are reached: atomic concepts or prime numbers. To help reinforce this analogy, I will refer to Leibniz's atomic concepts as *prime concepts* and to other concepts as *composite concepts*.

Representation of Concepts

For Leibniz the relation between prime and composite numbers and concepts is more than just an analogy: it is the basis for a knowledge representation system. Thus Leibniz assigns a prime number to every prime concept. Supposing, for the sake of the example, that *animal* and *rational* are prime, let us assign:

$$\text{animal} = a = 2, \quad \text{rational} = r = 3.$$

Then, supposing the correct definition of *man* is *rational animal*, we will assign to *man* the number 6:

$$\text{man} = m = ra = 3 \times 2 = 6.$$

Leibniz believed that every concept could be assigned exactly one number, which would reflect its analysis into prime concepts. In this he is directly in the Pythagorean tradition: everything is number, and intelligibility reduces to ratios (p. 21).[16]

Problem of Determining Primes

It is of course unlikely that Leibniz would have considered *rational* and *animal* to be prime. In fact he recognized that discovering the primes would be very difficult, and he gave only a few examples of concepts he thought might be prime, namely, *term, entity, existent, individual, I (ego)* (*C360*; *LLP* 51). He noted that for many purposes, it would be sufficient to have "relatively prime" concepts. For example, when we're doing geometry, we need be concerned only with the primes of geometry. Finding an adequate set of "atoms" remains a problem in contemporary knowledge representation languages.

Representation of Property Sets

Leibniz's representation becomes more familiar if we realize that he is using numbers as a way of representing finite sets of properties. Thus, if

[16] It is interesting that gematria is based on addition, whereas Leibniz' system is based on multiplication. This is an important difference. Numbers can be decomposed into sums in many different ways, and so gematria finds hidden connections among words that add up to the same quantity. On the other hand, numbers have a unique prime decomposition (into a product), which corresponds to a unique analysis into fundamental ideas.

A and R represent the properties *animal* and *rational*, then the essential attributes of *man* are represented by the set {R, A}. Leibniz saw that if each atomic concept were assigned a prime number, then every set of properties would have a single number representing it, for $ra = ar$, just as {R, A} = {A, R}. In modern AI programming we use linked lists for the same thing. Leibniz used numbers because they were the symbolic structures with which he was most familiar.[17]

Property sets (usually called property lists) are still widely used in AI as a representation for concepts. We will see shortly that Leibniz implemented inference by operations on these property sets that are also still in use.

Intension versus Extension

Lest you get confused, I must point out that there are two ways that concepts can be represented as sets, by *intension* or by *extension*.[18] The extensional approach is most familiar. The extension of a concept is the set of all the individuals to which that concept applies, that is, the individuals over which it *extends*. Thus the extension of the concept *man* includes Socrates, Hypatia, Leibniz, etc. We may write:

$$E(\text{man}) = \{\text{Socrates}, \text{Hypatia}, \text{Leibniz}, \ldots\}$$

On the other hand, the intension of a concept is the set of attributes possessed by that concept, which we may think of as the meaning or sense *intended* when we use the concept. Sometimes the intension is taken to include all attributes, so we may write:

$$\begin{aligned} I(\text{man}) \quad = \quad & \{\text{rational, animal, bipedal, language-using, primate,} \\ & \text{tool-making, mortal, } \ldots\} \end{aligned}$$

Other times, the intension is taken to be just the *essential* attributes:

$$I(\text{man}) = \{\text{rational}, \text{animal}\}$$

This is the sense in which we will use it, since then the intension is a *finite* set of attributes.

[17] It is interesting that when another mathematician, Gödel, needed a simple data structure in the days before computer programming, he also fell back upon the unique factorization theorem (see p. 284).

[18] The term 'intension' (with an 's'), which is under consideration here, must be carefully distinguished from 'intention' (with a 't'), which was discussed in Section 3.2. Joseph (*IL*, Ch. 6) has a good discussion of intension and extension.

Leibniz experimented with both the intensional and extensional approaches, but finally settled on the intensional. We will see (p. 127) that with Boole, formal logic took a definite turn in the extensional direction, and that the intensional logic was largely abandoned as unworkable. This may have been necessary for progress in logic at that time, but there has been a recent return to intensional representation for the same reason that Leibniz preferred it. This reason is that most concepts have an infinite extension (consider *man*). While mathematics is quite capable of handling infinite sets, computers are not.[19] Therefore, we cannot implement inference by performing operations on the infinite extensions of the concepts. On the other hand, intensions are finite, and the appropriate set operations are easy to implement (see also p. 121).

4.3.3 Computational Approach to Inference

Representation of Simple Propositions

By representing concepts intensionally in terms of prime concepts, Leibniz was able to explicate the meaning of propositions and define computational processes to determine their truth. Consider the proposition 'All S are P'. This means that each thing having the property S also has the property P. Therefore, the property P must be a part of the property S. That is, the prime concepts constituting P must be among those that constitute S, or $I(P) \subseteq I(S)$ (see above). In terms of Leibniz' numeric representation, the number representing P must evenly divide the number representing S; we may write $P \mid S$. Consider our previous example (p. 118): we know that all men are rational, since $r \mid m$, that is, $3 \mid 6$. Thus, if we know the correct numbers (definitions) of the concepts S and P, then we can decide the truth of 'all S are P' by a process of calculation.

Example Use of the Calculus

Our previous example, deciding the rationality of man by determining if r divides m, may seem pointless. Of course r divides m, because we defined $m = ra$. The value of the method may be clearer if we imagine that after many years of scientific and philosophical analysis the proper definitions of many concepts have been determined and collected into a philosophical dictionary. Since this would be the cumulative result of investigations in many sciences and many "layers" of definition, the implications of definitions

[19] More accurately, computers can deal with infinite sets, but only if they are represented intensionally; see MacLennan (*FP*, Ch. 7).

Set Operations on Extensions and Intensions

There is an interesting duality between set operations on the extensions and intensions of concepts. Consider the compound category PQ of things that are both P and Q. Clearly, the set of individuals that are both P and Q is the intersection of those that are P and those that are Q:

$$E(PQ) = E(P) \cap E(Q)$$

On the other hand, the properties of PQ things includes both the properties of P things and the properties of Q things. Thus:

$$I(PQ) = I(P) \cup I(Q)$$

Taking the union of the intensions is the same as taking the intersection of the extensions. It's also easy to see that all P are Q if the extension of P is contained in that of Q: $E(P) \subseteq E(Q)$. On the other hand all P are Q if the property of being Q belongs to everything that is P; thus the properties constituting Q are among those constituting P: $I(Q) \subseteq I(P)$. Unfortunately, this duality between the extensional and intensional operations breaks down when we consider the "sum" of concepts, whose extension is given by the union of the extensions:

$$E(P + Q) = E(P) \cup E(Q)$$

The intersection of the intensions of P and Q gives the genus to which P and Q both belong, but the extension of the genus may be more than the union of the extensions if P and Q do not exhaust this genus. However, in the case of dichotomous classification, duality is preserved.

would not be at all obvious. Thus, if we wanted to decide if chlorine is an oxidizing agent, we might look up these terms in our dictionary and find:

$$\text{chlorine} = 111546435 \text{ and oxidizing agent} = 255255$$

Dividing the first by the second we get 437 with no remainder, and thus know that chlorine is an oxidizer. Although this conclusion was implicit in the hierarchy of definitions, it was not apparent; the calculation has made it explicit.

For another example, suppose that we want to know whether chlorine is a metal. We look up metal = 36890, and divide into chlorine's number to get 3023.758... Hence, we conclude that it's not true that chlorine is a metal.

Representation of More Complex Inference

Leibniz's logical calculus is intended to support all the traditional moods and figures of the syllogism (pp. 52 and 79). By defining each kind of proposition computationally, he is able to explain the validity of syllogistic reasoning. For example, the validity of:

$$\begin{array}{c} \text{All } M \text{ is } P \\ \text{All } S \text{ is } M \\ \hline \text{All } S \text{ is } P \end{array}$$

simply follows from the fact that if P divides M and M divides S, then P divides S. In terms of their intensions, if $I(P) \subseteq I(M)$ and $I(M) \subseteq I(S)$, then $I(P) \subseteq I(S)$ (recall p. 121).

Representation of E, I, O Propositions

We have seen how the universal affirmative (A)[20] proposition is expressed numerically: all S is P if and only if P divides S. This shows immediately how to express a particular negative (O) proposition, since it's just the denial of the A. Thus some S is not P if and only if P does not divide S. In set terms, $I(P) \not\subseteq I(S)$ (p. 121). Leibniz had more difficulty with E and I. He explained the particular affirmative (I) proposition by saying that it means the predicate is contained in some species of the subject. Thus, some S is P if and only if for some X, P divides SX. In set terms, for some X, $I(P) \subseteq I(S) \cup I(X)$. The trouble is that there is always such an X, for example take $X = P$. This problem ultimately led Leibniz to abandon his representation of concepts by single numbers and to replace it by a representation in terms of pairs of numbers, which worked.

"Calculemus"

Leibniz had high hopes for his logical calculi. Once the (admittedly difficult) process of conceptual analysis had been completed, philosophical and

[20]The forms A, E, I and O are defined on p. 79.

scientific issues would be rationally decidable — in this case, literally by ratios. In a famous quotation:

> Then, in case of a difference of opinion, no discussion between two philosophers will be any longer necessary, as (it is not) between two calculators. It will rather be enough for them to take pen in hand, set themselves to the abacus, and (if it so pleases, at the invitation of a friend) say to one another: "*Calculemus*! [Let us calculate!]" (Leibniz, quoted in Bocheński, *HFL*, 38.11, p. 275)

Rationalist Optimism

This vision may now seem hopelessly naive and optimistic, but its possibility is implicit in the rationalist tradition to which Leibniz was heir. Coudert (*L&K*, p. 155) observes that his belief in progress and the perfectibility of humanity was grounded in his mystical, occult, and magical beliefs, for "The belief in the power and perspicuity of man arose in part from gnostic sources — from alchemy, Hermeticism, Renaissance Neoplatonism, and the Kabbalah" (Coudert, *L&K*, p. 155). Henceforth, "The denigration of reason and exaltation of faith so prevalent during the Reformation was reversed in the eighteenth century age of Enlightenment" (Coudert, *L&K*, p. 155).

The Mechanization of Reasoning

Finally, it should be noted that Leibniz had at his disposal all of the means necessary for the mechanization of reasoning, at least in principle. He had shown how logic could be reduced to a process of numerical calculation. Further, building on the pioneering projects of Schickardt and Pascal, Leibniz had constructed a mechanical calculator capable of multiplication and division.[21] Thus he had a machine that could — in principle — carry out the computations necessary to implement reasoning. We say "in principal" because in fact the capabilities of Leibniz's calculator were inadequate to handle the large numbers that would result from an actual implementation of his calculus. Indeed, the prime factorization of large numbers even taxes modern supercomputers (although divisibility tests are efficient). Of course, now we wouldn't use numbers at all; we would use some more efficient representation of the property sets. In fact, no one actually constructed a logic machine till Jevons' in 1869, and then it was constructed along very different principles (Section 4.5). Nevertheless we can see that Leibniz had all

[21] The first mechanical calculating machines (other than aids such as abaci and slide rules) were invented in 1623 by Wilhelm Schickardt (1592–1635) and in 1642 by Blaise Pascal (1623–1662). Leibniz constructed in 1671 a calculator that improved on Pascal's design by performing multiplication and division.

the components of a knowledge representation language and a mechanized inferential process to go with it. Nowadays we would call it an expert system.

4.3.4 Epistemological Implications

Scientific Facts are Necessary and Analytic

Socrates and Plato said that true knowledge (*epistēmē*), as opposed to right opinion, is knowledge of the eternal forms (Section 2.4). Aristotle said that scientific knowledge (*epistēmē*) is knowledge of universals (Section 2.5.2). This is a view that characterizes *rationalism* (p. 45), and Leibniz is a creature of this tradition. Again, Aristotle had said that every concept has a correct definition in terms of its essential attributes. For Leibniz this means that every concept has a unique prime decomposition, which we may discover by logical analysis. Hence, all scientific truths are ultimately established by analyzing the concepts involved; in technical terms, they are *analytic*:

> A statement is an analytic truth if and only if the concept of the predicate is contained in the concept of the subject… (Flew, *DP*, s.v. 'analytic')

But further, since there is only one correct analysis, and it depends only on the concepts involved, not on circumstances, these truths cannot be otherwise. In technical terms, they are *necessary*:

> [A] proposition is *necessary* if its truth is certifiable on *a priori* grounds, or on purely logical grounds. Necessity is thus, as it were, a stronger kind of truth, to be distinguished from the *contingent truth* of a proposition which might have been otherwise. (Runes, *Dict.*, s.v. 'necessary')

In summary, scientific knowledge — true knowledge — is necessary and analytic.

The Irrationality of the Contingent

We've seen how Leibniz's theory explains universal truths, such as 'all men are rational', but how can it account for particular truths, such as 'Socrates is a man' or 'Socrates died in 399 BCE'? Seemingly the "definition" of an individual contains an infinite number of properties, such as 'is a man', 'died in 399', 'had a snub nose', etc. It is implausible to suppose that they can be derived from a finite number of essential attributes. For this reason Leibniz thought that the intensions of particulars, such as individuals and historical events, comprised an infinity of prime concepts. Thus they were represented by infinite numbers. Finite intelligences such as ours could at best hope to accomplish a partial analysis of such numbers, eventually discovering more and more of the primes, but never being able to grasp them all. Leibniz believed that only the infinite mind of God could grasp such an infinite product of primes. Thus historical, *contingent* facts remain for us beyond the pale of science, and hence ultimately are irrational. But God is able to see how even these facts are necessary, and in fact analytic. Leibniz's rationalism is nothing if not comprehensive.

Summary

Leibniz designed the first workable knowledge representation language; it was based on intensions represented as property sets implemented by products of prime numbers. Leibniz also showed how inferential processes could be represented by operations on these property sets and how these operations could be reduced to calculation. Since calculation had already been mechanized, Leibniz had demonstrated in principle the mechanization of reasoning. Finally, Leibniz added additional support to the supposition that rational knowledge could be expressed in finite combinations of certain unanalyzable terms. In this he acknowledged the *irrationality* (infinite rational analysis) of the concrete, the particular, the individual and the historical.[22]

4.4 Boole: Symbolic Logic

> It cannot but be admitted that our views of the science of Logic must materially influence, perhaps mainly determine, our opinions upon the nature of the intellectual faculties.
>
> — Boole (*ILT*, p. 22)

> Boole resembles Aristotle both in point of originality and fruitfulness; indeed it is hard to name another logician, besides Frege, who has possessed these qualities to the same degree, after the founder.
>
> — Bocheński (*HFL*, p. 298)

[22]It should not be supposed that we've given even an overview of Leibniz's epistemological theories. For example, we've made no mention of his *monadology* or his *lex continui*, both of which are relevant to his theory of knowledge. Here we've restricted the discussion to his logical calculi and methods of knowledge representation and inference.

Linear Operator Calculus

An operator is a function that operates on other functions (such as differentiation and integration). An operator L is *linear* if it has the properties: $L(af) = aLf$ and $L(f + g) = Lf + Lg$. For example, for the derivative, $D(af) = aDf$ and $D(f + g) = Df + Dg$. Conversely, if L and L' are two linear operators, then we may write $(L + L')f = Lf + L'f$. Notice that application of a linear operator is very much like multiplication; it has similar formal properties. This formal relationship is exploited in the calculus of linear operators. For example, one might take the logarithm or exponential *of an operator* and expand it as a formal power series to achieve some end. Thus Boole shows $f(x + 1) = e^D f(x)$ by the formal expansion

$$e^D = 1 + D + \frac{D^2}{2} + \frac{D^3}{3} + \cdots$$

where D is the differentiation operator.

4.4.1 Background

George Boole: 1815–1864

Leibniz's goal of a calculus of reasoning was finally achieved by George Boole about 1854.[23] Although there were a number of contemporary efforts to produce such a calculus (by DeMorgan and others), it was Boole's that was most successful, and that determined the future direction of logic. Perhaps Boole's success resulted in part from his earlier use and development of the linear operator calculus (p. 126); this may have provided practical experience in tailoring a calculus to a specific purpose. It may have also led to the very mathematical (almost arithmetic) approach which he adopted, and for which he was criticized by later logicians (such as Jevons, Section 4.5).

The Laws of Thought

It should be noted that Boole's goal was not simply the development of a logical calculus; it is far more ambitious. His ideas are most fully developed in a book called *An Investigation of the Laws of Thought*, in which he says

[23] The principal source for this section is Boole (*ILT*).

that his purpose is

> to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolical language of a Calculus... (Boole, *ILT*, p. 1)

Thus his goal is a calculus that captures the mechanism of thought. To the extent that this can accomplished, he then will have reduced thinking (at least reasoning) to a mechanical process — ripe for implementation by a machine. This was in fact accomplished by Jevons about a decade later (Section 4.5).

Thought Manifested in Language

How could Boole hope to discover the laws of thought? The operation of the mind is not directly visible. However, he observed that "Language is an instrument of human reason, and not merely a medium for the expression of thought" (Boole, *ILT*, p. 24). We've seen that this identification of language and thought was generally accepted in the Western tradition. Boole claimed in addition that we can direct our investigation to the rules by which words are manipulated, since by "studying the laws of signs, we are in effect studying the manifested laws of reasoning" (Boole, *ILT*, p. 24). Conversely, of course, if there are forms of reasoning that do not manifest themselves verbally, then these will not be covered by Boole's laws. The possibility of nonverbal thought has been systematically ignored throughout much of Western intellectual history, however, and in this Boole is no exception.

Boole develops two logical systems: a class logic and a propositional logic. We will explore both of them briefly.

4.4.2 Class Logic

The class logic uses literal symbols $(x, y, \ldots)$ for classes and the symbols $+$, $-$ and $\times$ to represent operations on classes. The only relation among classes considered by Boole is identity $(=)$.

Extensional Interpretation

Boole consistently interprets terms extensionally (p. 119). Thus a term represents the class of individuals named by that term. In fact, Boole's class logic is essentially set theory. It may have been that this extensional approach was necessary for progress in logic at that time. For example, an extensional logic avoids the problem of identifying prime concepts, since the basic elements are individuals rather than atomic properties (recall p. 118). In any case Boole's thorough-going extensionism worked, and most logics since his time have been extensional. However, as noted previously, there has recently been renewed interest in intensional logic, since extensional logic is harder to implement on computers. Boole avoided this problem by ignoring the *mechanism* of reasoning.

Reason for Renewed Interest

As we saw (p. 120), while most concepts have infinite extensions, they have finite, often quite small, intensions. So long as logic was just a tool for mathematical analysis, the simpler, extensional approach could be used, because there was no need to actually manipulate the extensions. However, for artificial intelligence, knowledge structures must be representable in the computer's finite memory. This made the intensional approach more attractive, because a concept could be represented by a "property list" — a list of its defining properties.

Product of Classes

Boole takes a term (e.g., x or y) to represent a fundamental operation of thought: the selection of a class of things out of a wider class, or, as we may say, the focusing of the attention, which brings some things into the foreground, leaving the rest in the background. The mind can narrow its focus by successive selections. For example, if y represents the class of sheep and x represents the class of white things, then the product xy represents the result of first selecting the sheep from the universe, and then selecting the white things from the sheep. In other words, xy represents the white sheep. This process can be continued, for we may focus on the horned things (z) among the white sheep, zxy, that is, the class of horned white sheep. In modern terms, Boole's product xy is the intersection of the classes x and y, $x \cap y$.

Boole argues that his product is *commutative*, $xy = yx$, since the result of selecting all the xs from all the ys is the same as the result of selecting all the ys from all the xs. He gives examples (from Milton's poetry) to show that even in natural language, modifiers can be rearranged. Nevertheless, it is an interesting question whether focusing on sheep, and then narrowing the focus to white things, leaves the mind in the same state as focusing on white things and then narrowing to sheep. Are white sheep the same as ovine white things?[24] We may agree with Boole that they are *extensionally identical* — they select the same set individuals from the universe. The two expressions have the same *reference*, but it is not obvious whether they have the same *sense* or *meaning* (*intension*; recall p. 74, n. 16).

Idempotency

The same reasoning leads Boole to conclude that successive selection of the same class has no effect. For example, selecting the class of white things

[24] 'Ovine' denotes the property of being a sheep, or sheeplike.

from the class of white things is still just the class of white things, in symbolic form, $xx = x$. In the terminology of modern algebra, we say that the Boolean product is *idempotent* (the same power or efficacy, i.e., xx has the same efficacy as x). He admits that we sometimes repeat a word for emphasis (e.g., Burns' "red red rose") but "neither in strict reasoning nor in exact discourse is there any ground for such practice" (Boole, *ILT*, p. 32). On the other hand, it is apparent that the mental state resulting from hearing "a red rose" differs from that resulting from "a red red rose"; the red is more intense in the latter case. So here again, we have a divergence between sense (or meaning taken broadly) and reference.

It turns out that idempotency, which we may abbreviate $x^2 = x$, is one of the most characteristic features of Boole's system. In a negative sense, it is what distinguishes his algebra from everyday algebra, since, as he observes, the only numbers for which $x^2 = x$ are 0 and 1. Indeed, the idempotent property can be taken as the defining property of a *Boolean ring* (Halmos, *LBA*, p. 1). In a positive sense, idempotency allows important theorems to be proved, as we'll see shortly.

Sum of Classes

Another operation of thought, according to Boole, is the *sum* of classes, $x + y$. For example, if x is trees and y is minerals, then $x + y$ is the class of trees and minerals. He observes that in English this operation may be signified by either 'and' or 'or'; for example, in both "Italians and Germans may join" and "Members must be Italians or Germans" the sum of the classes is indicated. This shows that there is not a direct relation between 'and' and 'or' in natural language and in logic.

Boole argues that the sum of classes is commutative, $x + y = y + x$, and that the product *distributes* over the sum,

$$x(y + z) = xy + xz.$$

For example, "European men and women" refers to the same individuals as "European men and European women."

Sum of Classes as Disjoint Union

Boole argues that $x + y$ makes sense only if x and y are disjoint classes, so we can form the sum of *minerals* and *trees*, but not of *philosophers* and *scientists*, because some people are both. In modern terms, Boole's '+' represents a *disjoint union* (or *exclusive* 'or'), which he claims is a more basic operation of thought than the inclusive operation, which allows overlap. It's certainly true that in everyday speech we often take 'or' exclusively, unless the possibility of overlap is explicitly indicated, for example, "philosophers or scientists or both." Nevertheless, later logicians considered this decision to be one of Boole's mistakes, and contemporary symbolic logic and set theory take the inclusive operations to be more basic.

Advantages of Exclusive Definition

This is a case, however, where Boole's logic may have been an improvement on its successors, for there are advantages to the Boolean definition. As we'll see, the exclusive definition facilitates algebraic manipulation and simplifies the solution of logical equations, both of which were among Boole's explicit goals. It is significant that in the further development of abstract algebra, as well as in its application in programming languages, Boole's disjoint union has been found more useful than the traditional (inclusive) union (MacLane & Birkhoff, *Alg*; MacLennan, *FP*, Sec. 5.2).

Difference of Classes

Boole observes that a sum operation immediately suggests a difference operation, and so he proposes $x - y$ to represent expressions such as 'Europeans except Germans'. He says that $x - y$ makes sense only if the class x includes the class y (as in the example). This again is different from the modern set difference, but the Boolean definition may be better for algebraic manipulation.[25] The properties satisfied by the difference include distributivity, $z(x - y) = zx - zy$, and the very important *transposition* property,

$$x = y + z \quad \text{if and only if} \quad x - y = z,$$

which facilitates the solution of equations in logic just as it does in elementary algebra.

Negative Classes and Formalism

Boole says "it is indifferent for all the *essential* purposes of logic" whether we write $x - y$ or $-y + x$, but what does the expression '$-y$', the "negative" of a class, mean? It is not the difference of the universe and y (which we will see Boole writes $1 - y$), because when we add it to x it takes the elements of y *out* of x. We could attach significance to $-y$ by inventing some class for it to name, perhaps a class of things like the elements of y but having some kind of "negative existence," a sort of "anti-matter" to cancel the elements of y. Boole has a simpler solution: he simply treats $-y$ as a *formal expression* that has no meaning apart from its use in formulas such as $-y + x$. In other words, $-y$ has significance only when it is transformable into a context such as $x - y$. This is true *formalism* — the symbols have significance only through their relation with other symbols, not through any intrinsic meaning — and Boole seems to be the first logician to have recognized its power.

[25] In modern set theory $x - y$ is the set of all things that are in x but not in y, so y may have members not in x.

> They who are acquainted with the present state of the theory of Symbolic Algebra, are aware, that the validity of the processes of analysis does not depend upon the interpretation of the symbols which are employed, but solely upon the laws of their combination. (Boole, quoted in Bocheński, *HFL*, § 38.17)

He goes on to observe that we may impose any interpretation we like on the symbols, so long as their algebraic properties are preserved; we'll see examples shortly (Section 4.4.3).

Extreme Classes

Boole uses the terms 0 and 1 for the *extreme classes* "nothing" and "everything." They have the obvious algebraic properties

$$0x = x0 = 0, \quad 0 + x = x + 0 = x, \quad 1x = x1 = x,$$

which are useful in solving equations. One important use of 0 is to represent the mutual exclusiveness of classes, thus $xy = 0$ means that nothing is both x and y. Similarly 1 is used to form the complement of a class; thus $1 - x$ is the class of all things that are not x.

Identity of Classes

Boole uses only one relation between classes: *identity*, represented by '='. Most notable is that he has no operation for class inclusion analogous to the modern subset relation, $x \subseteq y$, although it had been invented 30 years before (Bocheński, *HFL*, pp. 303–305). Thus he has to express "all x are y," which we write $x \subseteq y$, by the circumlocution $x(1 - y) = 0$, that is there are no xs that are not y. Although Boole has been criticized for his lack of an inclusion relation, he may again be smarter than his critics, since equational reasoning is easier and more powerful than relational reasoning. Most people find it easier to reason about equalities than about inequalities.

Example of Proof: Law of Contradiction

To give a bit of the flavor of Boole's algebraic logic, I will present one of his proofs. First he shows that he can "prove" the Law of Contradiction — that something cannot be both x and non-x — which had been accepted as an axiom since Aristotle (*Met.* 1005b19–22). Boole begins with idempotency, $x^2 = x$; applies transposition, $x - x^2 = 0$; and then undistributes the product, $x(1 - x) = 0$, that is, nothing is both x and non-x. Of course, he hasn't *really* proved the Law, since his proof depends on it; for example, by stating $x^2 = x$ he means to exclude $x^2 \neq x$. What he has done is show how the Law of Contradiction is related to other properties, such as idempotency. This is important but not so astonishing.

Importance of Second Degree Equations

With regard to the fundamental equation $x^2 = x$, Boole (*ILT*, p. 50) makes a thought-provoking claim:

> Thus it is a consequence of the fact that the fundamental equation of thought is of the second degree, that we perform the operation of analysis and classification, by division into pairs of opposites, or, as it is technically said, by *dichotomy.*

Had it been otherwise, "the whole procedure of the understanding would have been different." In particular, if the equation had been of the third degree, then *trichotomy* would be the basic procedure of analysis, but he says that the nature of this is "impossible for us, with our existing faculties, adequately to conceive." In fact, the equation $x^3 = x$ can be written

$$x(1-x)(1+x) = 0,$$

which holds if x is restricted to the values $\{-1, 0, 1\}$, and suggests a three-valued logic. Finally, if we suppose that repetition always has an effect, then $x^p = x$ will not be true in general for any p, which suggests that no p-fold division will be adequate. I will leave the exploration of these ideas as an exercise for the reader!

4.4.3 Propositional Logic

Terms

I have mentioned before that, although a propositional logic was investigated by the Megarian-Stoic logicians, a shift from a logic of classes to a logic of propositions was an important nineteenth century development. Boole's system illustrates this change, for it is simultaneously a class logic and a propositional logic. In the class logic a term such as x represents the set of *objects* belonging to the class. In the propositional logic a term represents the set of *situations* in which the proposition is true. For example, if x is the class of situations in which it is day, then it represents the proposition 'it is day'.

Propositional Product and Sum

Terms representing propositions can be combined in the same way as terms representing classes. For example if y is the proposition 'it is light', then xy is the proposition 'it is day and it is light', because xy is the class of situations in which it is both day and light. In terms of mental operations, we focus on the situations in which it is day (x), and then from those we select the situations in which it is light (y). Similarly, $x+y$ is the proposition that it is day *or* it is light (but not both — exclusive 'or').

Extreme Propositions

Just as we have the *extreme classes* 0 and 1 ("nothing" and "everything") in the class logic, so we have the *extreme propositions* 0 and 1 in the propositional logic, in which 0 is the proposition that is not true in any situation

— that is, it's never true — and, conversely, 1 is true in all situations. Since the proposition 0 is always false, and 1 is always true, they are called *truth values*. So we can say that Boole's logic has two truth values 0 and 1.[26]

Boolean Logic and Computers

It perhaps not surprising that Boolean algebra has found application in the design of digital computers and other digital logic systems. The truth values 0 and 1 can be represented by low and high voltages or other physical quantities, and "logic gates" ('not', 'and', inclusive and exclusive 'or') are implemented by simple circuits. This was first described in the MS thesis of Claude Shannon (SARSC), who is best known for his later development of information theory. This is a direct result of the *formality* of Boole's system: any phenomena that have the same *form*, that is, that obey the same algebraic laws (idempotency, commutativity, distributivity, etc.), can be analyzed with Boolean algebra and can implement Boolean logic. (In Section 4.5 I'll describe a *mechanical* implementation.)

4.4.4 Probabilistic Logic

Formal Identity to the Field **2**

Boole's operations, $+$, $-$, $\times$, are obviously very similar to the familiar arithmetic operations. As I've noted, a principal difference is that his product is idempotent, $xx = x$, whereas the familiar product is not. Boole observes that his algebra is the same as the familiar algebra, but restricted to the two numbers 0 and 1. In particular, if we do all the arithmetic *modulo 2*, so that $1+1=0$, then the Boolean operations are the same as modulo 2 arithmetic:

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

In mathematical terms, this is arithmetic *over the field* **2** (where $\mathbf{2} = \{0, 1\}$.) To a large extent, algebra over *any* field is the same, and so the mathematics is like highschool algebra. As long as we are careful about peculiar properties such as idempotency, we can solve equations in the same way no matter whether the variables refer to integers, bits, classes, truth values, or integers modulo some number. This is the value of *formality*.

Probability Calculus

Boole himself provides a convincing example of the power of a formal system, for he shows that his algebra can also be interpreted as a *calculus*

[26] In computer science, the truth values *true* and *false*, which are equivalent to the bits 1 and 0, are commonly called *Boolean values*.

of probability. Instead of interpreting a term x as the class of situations in which a proposition is true, we interpret it as the *probability* of that proposition being true, more carefully, as the ratio of the number of (equally likely) situations in which it is true to the total number of (equally likely) situations. Then, if x and y represent two elementary (independent) events, the probability of both occuring will be xy and the probability of either (but not both) occuring will be $x+y-xy$. As expected, 0 represents the impossible, 1 represents the certain, and $1-x$ represents the probability of x not occuring. Notice that all Boole has done is to expand the domain of the variables from the two-element set $\{0,1\}$ to the continuous interval $[0,1]$. As a result, most of the mathematics goes through unchanged (although certain properties, such as idempotency, no longer hold).

4.4.5 Summary

Boole constructed the first really successful mathematical logic. Contributing to this were the wise choice of an extensional viewpoint — which allowed him to circumvent many epistemological problems — and a deep understanding of the power of *algebra* — which may have come from his earlier use of formal methods with differential and difference equations (see p. 135). His algebra is simultaneously a logic of classes, propositions and probabilities. Although Boole's logic is no longer used in its original form, Boolean algebra is still widely used in digital circuit design, and his logic formed the foundation for modern symbolic logic and set theory.

We also find in Boole the first clear statement of the idea of *formality*: the separation of the symbolic rules from their interpretation. This paves the way for the computer implementation of formal processes (including deduction), since it shows that they can be implemented in any way, so long as the formal properties are preserved. In fact, the first logic *machine* followed Boole's work by only 15 years (see the next section). One corollary is that *if* the "laws of thought" are truly formal, then they can be implemented in a computer as well as in a brain, and so a computer can think in the same sense as can a person. This is the issue addressed in Searle's well-known "Chinese Room Argument" against AI.

Logical Taylor Series

A fascinating demonstration of algebraic manipulation in Boolean algebra is the use of Taylor's theorem to expand a *Boolean* function (Boole, *ILT*, pp. 72–3, note). We begin with the usual Taylor (actually, Maclauren) expansion:

$$f(x) = f(0) + f'(0)\frac{x}{1} + f''(0)\frac{x^2}{1 \cdot 2} + f'''(0)\frac{x^3}{1 \cdot 2 \cdot 3} + \cdots.$$

But the Boolean product is idempotent, $x = x^2 = x^3 = \cdots$, so the expansion reduces to

$$f(x) = f(0) + \left[\frac{f'(0)}{1} + \frac{f''(0)}{1 \cdot 2} + \frac{f'''(0)}{1 \cdot 2 \cdot 3} + \cdots\right]x. \tag{4.1}$$

For the special case $x = 1$ this becomes

$$f(1) = f(0) + \frac{f'(0)}{1} + \frac{f''(0)}{1 \cdot 2} + \frac{f'''(0)}{1 \cdot 2 \cdot 3} + \cdots,$$

and so

$$f(1) - f(0) = \frac{f'(0)}{1} + \frac{f''(0)}{1 \cdot 2} + \frac{f'''(0)}{1 \cdot 2 \cdot 3} + \cdots.$$

The right-hand side of this equation is the bracketed expression in Eq. 4.1, so replace the bracketed expression by $f(1) - f(0)$ to get

$$f(x) = f(0) + [f(1) - f(0)]x.$$

This is the Maclauren expansion for Boolean functions; it can be easily rearranged into the more transparent form:

$$f(x) = f(1)x + f(0)(1 - x).$$

Thus we can express an arbitrary Boolean function in terms of its value on the two special values 0 and 1. Notice in particular that it is as true for classes as for propositions.

4.5 Jevons: Logic Machines

> As I awoke in the morning, the sun was shining brightly into my room. There was a consciousness on my mind that I was the discoverer of the true logic of the future. For a few minutes I felt such a delight such as one can seldom hope to feel.
>
> — Jevons (Mays & Henry, J&L)

> When contemplating the properties of this Alphabet I am often inclined to think that Pythagoras perceived the deep logical importance of duality; for while unity was the symbol of identity and harmony, he described the number two as the origin of contrasts, or the symbol of diversity, division and separation.
>
> — Jevons (*PS*, p. 95)

W. Stanley Jevons: 1835–1882

William Stanley Jevons was a versatile nineteenth century philosopher-scientist.[27] In addition to major contributions to economics, including a mathematical theory of economic utility and the use of statistical data in the analysis of economic trends, he conducted research in meteorology, developed a philosophy of science that was quite far ahead of its time (Jevons, *PS*), and wrote a widely used handbook of logic (Jevons, *ELL*). Here however we will be concerned with his development in 1869 of the first machine to implement deductive logic.

4.5.1 Combinatorial Logic

Laws of Thought

Jevons was lavish in his praise of Boole; for example: "Undoubtably Boole's life marks an era in the science of human reason." Nevertheless he pointed to several improvements that he made on Boole's system. For example, he reduced the axioms to three "Laws of Thought":

[27] Primary sources for Jevons are Jevons (*PS*, pp. 91–96, 104–114), Jevons (*ELL*, pp. 196–201) and Jevons (OMPLI). Secondary sources are Gardner (*LMD*, Ch. 5) Mays & Henry (J&L) and Nagel (Jevons, *PS*, Intro. to Dover Ed.).

Law of Identity:	$A = A$
Law of Contradiction:	$Aa = 0$
Law of Duality:	$A = AB \cdot\!|\!\cdot Ab$

Here we see several particulars of Jevons' notation. First he uses lowercase letters for the complements of classes, thus a is the class non-A. Second, he introduced the symbol '$\cdot|\cdot$' for the *inclusive* disjunction, or union, of two classes; this is generally considered an important technical advance over Boole's exclusive disjunction. In Jevon's notation $A \cdot | \cdot B$ means the class of things that are As or Bs or both; now we would write $A \cup B$.[28]

The Logical Alphabet

Another improvement claimed by Jevons was the development of a mechanical, combinatorial approach to deduction. This was based on the use of the *Logical Alphabet*, to which Jevons attached great significance:

> It holds in logical science a position the importance of which cannot be exaggerated, and as we proceed from logical to mathematical considerations, it will become apparent that there is a close connection between these combinations and the fundamental theorems of mathematical science. (Jevons, *PS*, p. 93)

The Logical Alphabet, for a given set of terms, is simply an enumeration of all possible conjunctions of those terms and their negations. Table 4.1 shows the 2-term, 3-term and 4-term Logical Alphabets. Clearly, if terms are assigned specific positions, then the 2^n combinations of the n-term Logical Alphabet are equivalent to the n-bit binary numbers.[29] Also notice the similarity to Lull's enumeration of combinations; the principal difference is that Jevons explicitly distinguishes a class and its complement, whereas Lull left that to the operator.

Dichotomy

Jevons (*PS*, pp. 702–704) relates the Logical Alphabet to the Tree of Porphyry (or Ramean Tree), which we have already considered (Section 3.1). The *summum genus* or universal class is divided into mutually-exclusive, disjoint classes A and non-A, or in Jevons' notation, A and a. Each of these may be further divided on the basis of some trait B, yielding classes AB, Ab, aB, ab, and so forth. I have already quoted (p. 136) his statement concerning the significance of binary division; he goes on to say:

[28]As did Boole, Jevons wrote the intersection of classes, $A \cap B$, as a product, AB, as mathematicians still sometimes do. Also recall that the Stoic-Megarian logicians had already noted the importance of the inclusive disjunction in propositional logic, but Jevons did not foresee the coming shift from class logic to propositional logic.

[29]Recall Leibnitz' invention of binary numbers (p. 115).

Table 4.1: Logical Alphabets for 2, 3 and 4 Terms

II	III	IV
AB	ABC	ABCD
Ab	ABc	ABCd
aB	AbC	ABcD
ab	Abc	ABcd
	aBC	AbCD
	aBc	AbCd
	abC	AbcD
	abc	Abcd
		aBCD
		aBCd
		aBcD
		aBcd
		abCD
		abCd
		abcD
		abcd

> The followers of Pythagoras may have shrouded their master's doctrines in mysterious and superstitious notions, but in many points these doctrines seem to have some basis in logical philosophy. (Jevons, *PS*, p. 95)

Or — I would add — to be the basis *of* logical philosophy.

The Indirect Method

Although Jevons acknowledges the value of Boole's mathematical logic, its limitation is that the solution of equations requires human intelligence. The main utility of the Logical Alphabet is that it allows deduction to be carried out mechanically. The basic approach, which he calls the *Indirect Method* (or *Indirect Deduction*) is one that had been used in the Middle Ages: enumerate all the possibilities and cross out those inconsistent with the premises (i.e., generate and test).

> The method of Indirect Deduction may be described as that which points out what a thing is, by showing that it cannot be anything else. (Jevons, *PS*, p. 81)

Gardner calls it a combination of Ramon Lull and Sherlock Holmes ("when you have eliminated the impossible, whatever remains, *however improbable*, must be the truth" — *The Sign of Four*).

I will illustrate the method by working through one of Jevons' examples. Consider the premises:

1. A must be either B or C;
2. B must be D;
3. C must be D.

For instance, we might have

$$\begin{array}{rcl} A & = & \text{organic substance,} \\ B & = & \text{vegetable substance,} \\ C & = & \text{animal substance,} \\ D & = & \text{consisting mainly of carbon, nitrogen and oxygen.} \end{array}$$

We start with the 4-term alphabet (Table 4.1, IV). Premise (1) is inconsistent with $AbcD$ and $Abcd$, since in these cases A is neither B nor C, so we cross these out.[30] Similarly premise (2) is inconsistent with $ABCd$ and $ABcd$ since

[30] It's apparent that the Indirect Method can be implemented by simple binary operations.

in these cases we have B but not D. Finally, premise (3) is inconsistent with $AbCd$, which leaves the combinations:

$$ABCD, \quad ABcD, \quad AbCD.$$

This gives the "complete solution" of the logical equations, that is, all of the possibilities that are consistent with the premises.

Disjunctive Normal Form

The Indirect Method can be understood in contemporary terms as follows. The logical alphabet represents the possibilities by a disjunction of all possible conjunctions of every term or its complement:

$$ABCD \cdot|\cdot ABCd \cdot|\cdot ABcD \cdot|\cdot \;\cdots\; \cdot|\cdot abcd.$$

The method proceeds by striking from this disjunction all conjunctions inconsistent with the premises, in this case yielding

$$ABCD \;\cdot|\cdot\; ABcD \;\cdot|\cdot\; AbCD,$$

which is a formula expressing the maximal result consistent with the premises. In modern terms this is a formula in *disjunctive normal form*, which is still widely used in artificial intelligence.

The Inverse Problem

The difficulty with Jevons' method is that the result of the analysis may be difficult to interpret. What are we to make of $ABCD$, $ABcD$, $AbCD$? If we look at it, we can see that in every case where A is present, so is D; therefore we can conclude all A are D, or all organic substances consist mainly of carbon, nitrogen and oxygen. This is presumably the intended conclusion, but Jevons' solution includes many others that are less interesting.

This process of converting a result to intelligible form is called by Jevons "the abstraction of indifferent circumstances" or "the inverse problem," and he acknowledges that "it is one of the most important operations in the whole sphere of reasoning." Unfortunately, his calculus is no help in this regard, and in fact aggravates the problem by enumerating all the possibilities.[31] The difficulty is that we want the *relevant* conclusion implicit in the complete solution, but the evaluation of relevance remains an unsolved problem in AI.

4.5.2 Logic Machines

The Logical Slate

We have seen how Jevons reduced deduction to a mechanical generate-and-test process; unfortunately, the mindlessness of the process makes it tedious

[31] Another limitation of Jevons' approach is that it was very difficult to handle *particular propositions*, such as "some animals are mammals."

and hence error prone, a problem of which Jevens was well aware:

> I have given much attention, therefore, to lessening both the manual and mental labour of the process, and I shall describe several devices which may be adopted for saving trouble and risk of mistake.(Jevons, *PS*, p. 96)

In fact we see him producing a series of devices that are progressively more automatic.

At first he simply automated the writing out of the Logical Alphabets, either by having copies printed in advance or by making stamps to generate them when needed. Later he invented the Logical Slate, which has an alphabet engraved on its surface so that chalk can be used to strike out combinations, and later erased.

The Logical Abacus

The Logical Slate is not very convenient for executing certain deductive processes that involve reintroducing previously canceled combinations (you'll see an example shortly). Therefore Jevons took another step towards automation with the Logical Abacus. This device has an inclined surface with four ledges that can hold thin wooden cards bearing the combinations of a 2-, 3- or 4-term alphabet (Fig. 4.2). In addition these cards have a pin for each term, but in one position for an uppercase term (A) and another for a lowercase (a). This allows a metal rule (*gnômôn*!) to be used to lift all the combinations having a common term (e.g., all the As or all the as).

> Thus we have represented the act of thought which separates the class A from what is not-A.(Jevons, *PS*, pp. 104–105)

In modern computer terms we would say that the rule selects all the terms with a given bit set to a specified value (0 or 1).

We'll work through a simple example, a syllogism of the form "all A are B, all B are C, therefore all A are C." First the premises must be expressed as logical equations:

$$A = AB, \quad B = BC.$$

To execute the deduction we begin with the complete 3-term alphabet on shelf 2, and then raise all the as to shelf 1, which leaves on shelf 2 the combinations containing A. From these remove the bs (since they are inconsistent with $A = AB$) and lower them to shelf 3. Now take the a combinations from shelf 1 and lower them back into shelf 2 (since the equation $A = AB$ isn't inconsistent with a). Now on shelf 2 we have the a and the AB combinations,

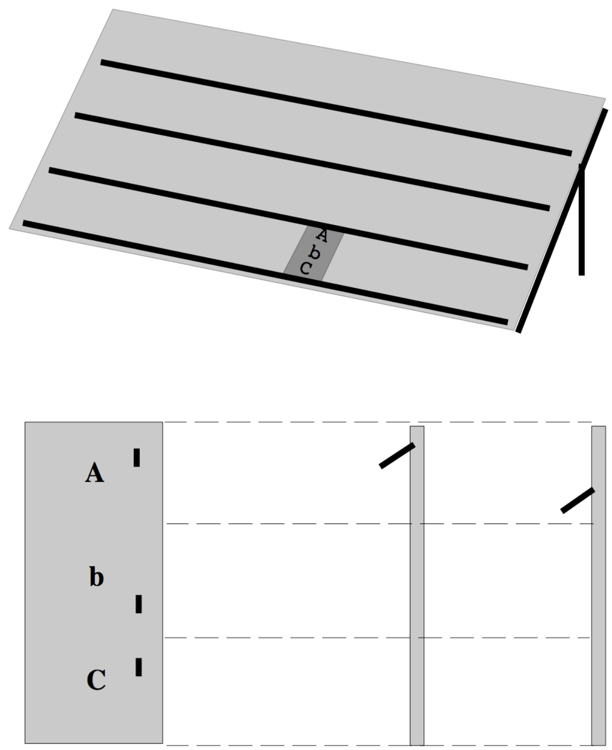


Figure 4.2: The Logical Abacus. The inclined board has four ledges, each capable of holding the cards representing the 3-term logical alphabet; one such card is shown on the lowest ledge. The lower part of the figure shows a card in detail. The face of the card is marked with the terms A, B, C is either their positive (A) or negative (a) form (lower left figure). In addition, each card has three pins driven into its face, in a higher position if the term is positive (lower middle figure), and in a lower position if negative (lower right figure). As described in the text, a metal rule can be used to manipulate the cards by means of their pins.

which represents "not A or both A and B," which is equivalent to the first premise. All that to "enter" the first equation into the device!

The second equation is of the same form, and is handled the same way: raise b, lower Bc, and relower b. The end result is that shelf 2 holds

$$2. \quad ABC, aBC, abC, abc,$$

which represent conditions consistent with the premises, and shelf 3 has

$$3. \quad ABc, AbC, Abc, aBc,$$

which represents those inconsistent with them. As usual, the "inverse problem" is difficult, but the expected result, all A are C, is among the combinations on shelf 2. This is easier to see if we raise the combinations containing A, in effect asking "What does A imply?" Only card ABC is selected, telling us A implies B, which we knew, and A implies C, which we (presumably) didn't know. You can see that although the process is mechanical, it is still quite complicated, both in execution and interpretation. Therefore Jevons automated again.

Other Computers

Jevons was not the only one designing computers in the mid-nineteenth century. He was aware of the "difference engines" designed by Babbage and Scheutz, but said that they were for *quantitative* reasoning, whereas his machine was for *qualitative* reasoning. He claims that the only other proposals for a *logic* machine are Swift's professors of Lagado [32] (p. 83), and Alfred Smee's 1851 proposal to construct

> a Relational machine and a Differential machine, the first of which would be a mechanical dictionary and the second a mode of comparing ideas ... (Jevons, *PS*, p. 107),

which, however, Jevons claims would have never worked.[33]

The Logical Machine

Figure 4.3 shows the appearance of Jevons' Logical Machine or Logical Piano, as he sometimes called it. It is remarkably modern in appearance (more so than early computers such as ENIAC), with a 21-key keyboard and a "display" showing a kind of spreadsheet of logical possibilities. After two unsuccessful attempts the machine was constructed in 1869 and demonstrated for the Fellows of the Royal Society on January 20, 1870 (Jevons, OMPLI; Jevons, *ELL*, p. 200; Jevons, *PS*, p. 108). I'll briefly describe its operation.

[32] Not Laputa, as Jevons, Bonner, Gardner and many others say.

[33] Gardner (*LMD*) describes many other ingenious logic devices.

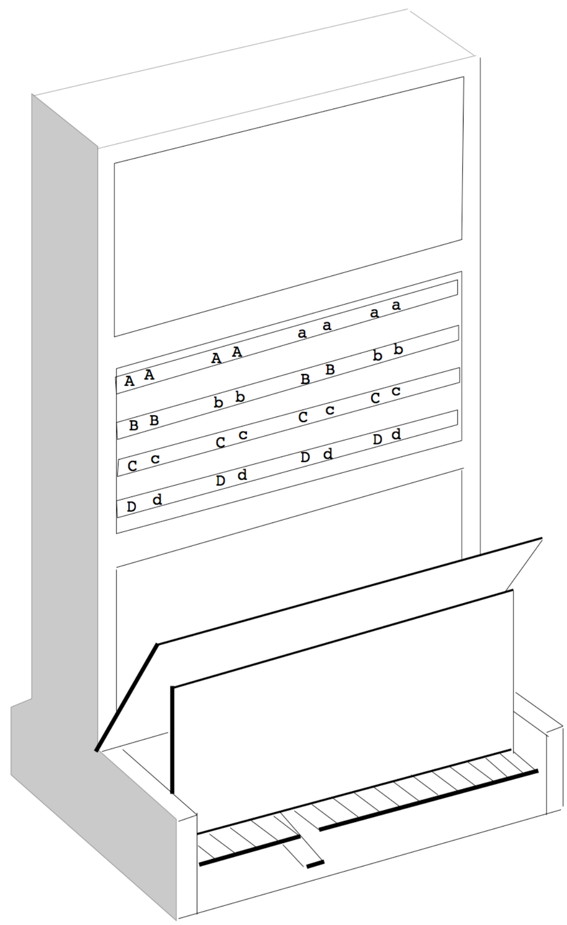


Figure 4.3: The Logical Machine. Shown with the keyboard open and one key depressed. The vertical "display" shows all combinations of a 4-term Logical Alphabet that are consistent with the equations entered so far. (Based on a drawing in Jevons, *PS*.)

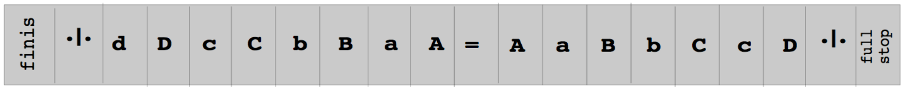


Figure 4.4: The Keyboard of the Logical Machine. The alphabet keys on the left are used for terms on the left-hand side of an equation, and similarly for those on the right. The central key is used for the '=' of the equation, and the 'full stop' key is pressed after each equation. The 'finis' key clears the machine so that a new set of equations can be entered. The '·|·' keys are used for disjunctions on either the left- or right-hand sides of the equations. (Based on a drawing in Jevons, *PS*.)

Pressing the Finis key on the far left of the keyboard (Fig. 4.4) signifies the completion of one problem and the start of another, so it resets the display to show all the combinations of a 4-term logical alphabet. (For smaller problems the unneeded terms must be ignored, a distinct inconvenience.) The equations are then typed on the keyboard just as they are written:

$$\begin{aligned} A &= AB. \\ B &= BC. \end{aligned}$$

The period character '.' represents the Full Stop key on the far right of the keyboard; it signals that a complete equation has been entered. The only complication is that there are two sets of letter keys, one for the terms on the right-hand side of the equation, another for those on the left. On the other hand, the equations are not limited in number or length, and can be quite complex in form:

$$\begin{aligned} A &= AB \cdot | \cdot AC. \\ B \cdot | \cdot C &= BD \cdot | \cdot CD. \end{aligned}$$

(There are some complications in equation entry, but I will pass over them.)

As each equation is entered the display is updated to show the combinations that are still consistent with the premises, so the user can watch the developing logical analysis of the problem. We can also ask the consequences of various logical situations. For example, typing Ab will show the combinations consistent with A and not B. Typing a Full Stop restores the machine to its previous state and allows a different combination to be tried.

Combinatorial explosion

Overall, the Logical Piano was a remarkable machine for its time. However, the Jevons' logic machine also illustrates the limitations of a combinatorial approach to inference. He apparently planned to build a 10-term machine, but abandoned the project when he calculated that it would occupy one entire wall of his study (Gardner, *LMD*, p. 100). (Note that the 10-term logical alphabet has 1024 combinations.)

4.5.3 Discussion

Jevons is typically a rather sober writer, but the metaphors fly fast when he describes the Logical Machine:

> After the Finis key has been used the machine represents a mind endowed with powers of thought, but wholly devoid of knowledge. . . . But when any proposition is worked upon the keys, the machine analyses and digests the meaning of it and becomes charged with the knowledge embodied in that proposition.(Jevons, *PS*, pp. 110–111)

I suppose we must also credit Jevons with the invention of "AI hype"! When we remind ourselves that Jevons' machine is just a device that implements some simple binary operations on 4-bit strings, it becomes apparent how easy it is to attribute mental qualities to a machine. To take another example:

> It cannot be asserted indeed that the machine entirely supersedes the agency of conscious thought; mental labour is required in interpreting the meaning of grammatical expressions, and in correctly impressing that meaning on the machine; it is further required in gathering the conclusion from the remaining combinations. Nevertheless the true process of logical inference is really accomplished in a purely mechanical manner.(Jevons, *PS*, p. 111)

Here we find Jevons encountering — but failing to recognize — what continues to be a central problem in the philosophy of artificial intelligence, the problem of *original versus derived intentionality* (see, e.g., Dennett, *IS*). The basic issue is this: we agree that language has (or can have) meaning when we use it, but can a proposition inside a computer have any meaning beyond the meaning *we* impose on it? Can a computer *really* understand something, or only *appear* to understand it? These are complex issues that will be faced in Volume 2.

Relevance of Logic to Common Sense

Finally we return to Jevons the sober scientist:

> I may remark that these mechanical devices are not likely to possess much practical utility. We do not require in common life to be constantly solving complex logical questions.(Jevons, *PS*, p. 112)

This is another timely question, which will also be addressed again in Volume 2: What is the role of logic in everyday human activity? Traditional AI and cognitive science have assumed that logic is a good model of human behavior. More recent evidence suggests that people (and other animals) are illogical in important — survival enhancing — ways. But first we must explore the final working out of the Pythagorean program.

Part II

The Triumph of the Discrete

Chapter 5

The Arithmetization of Geometry

5.1 Descartes: Geometry and Algebra

5.1.1 Arabian Mathematics

For "Is" and "Is-not" though with Rule and Line
and "Up-and-down" by Logic I define,
  Of all that one should care to fathom, I
Was never deep in anything but — Wine.

— Fitzgerald (*Ruba'iyat of Omar Khayyām*, 56)

For a while, when young, we frequented a master,
For a while we were contented with our mastery;
Behold the end of words:— what befell us?
We came like water and we went like wind.

— Omar Khayyām (Ouseley MS. 140:121, Bodleian Library)

On the Brink of Unification

Our goal is to understand the relation between the continuous and the discrete, and especially the roots of the historical preference for the discrete. In the previous chapter we saw the development of the idea that thought is *calculation* — discrete symbol manipulation — and that this calculation could be mechanized. The reduction of logic to discrete symbol manipulation is on the direct path to traditional AI, and we will return to that path later. In this chapter, however, we will look at the triumph of the discrete in mathematics and the reduction of change to ratios — this amounts to the completion of the task begun by Pythagoras, the arithmetization of geometry and nature. Since we are so embedded in the discrete viewpoint, it will be helpful to pause on the brink of the unification, at a time when the continuous and discrete still seemed irreconcilable and change seemed irrational. This will help us to view the relation between the continuous and the discrete from a more balanced perspective. Although this chapter is mainly concerned with developments in mathematics and science in the sixteenth and seventeenth centuries, it will be necessary to consider their historical antecedents.

The End of Learning in Europe

An important phase in the history of mathematics is its development at the hands of Hindu and Arabic scholars, so we should consider its background.[1] The rise of Christianity brought the decline of Pagan science. The early Christians on the whole did not value science, perhaps because of their other-worldly focus, and their expectation of the immanent end of the world. In any case, neglect turned to open persecution after Christianity was made the official religion of the Roman Empire by Constantine (272–337), and the persecution accelerated in the fourth century when Theodosius banned Paganism and ordered the destruction of its temples. In 391 Archbishop Theophilus — whom Gibbon described as "the perpetual enemy of peace and virtue; a bold, bad man, whose hands were alternately polluted with gold, and with blood" — led a mob of Christians to the temple of Serapis in Alexandria, where they destroyed the sculptures and the library; perhaps 300 000 manuscripts were burned.

Hypatia

The empire had become a dangerous place for Pagan scholars; well-known is the story of Hypatia (born, 365), the highly respected Alexandrian philosopher, mathematician and astronomer, known for her intelligence, character and beauty. Saint Cyril, nephew of Theophilus and then Bishop of Alexandria, was a specialist at inciting mobs to violence, especially against the Jews in Alexandria, and eventually he destroyed their synagogues and drove them from the city. In 415 he provoked a Christian mob against Hypatia and in Gibbon's words:

> On a fatal day, in the holy season of Lent, Hypatia was torn from her chariot, stripped naked, dragged to the church, and

[1] Sources include Kline (*MT*, Chs. 8–10), Ronan (*Science*, Chs. 4, 5), Chuvin (*CLP*, Chs. 5, 6, 10) and Canfora (*VL*).

> inhumanly butchered by the hands of Peter the reader and a troop of savage and merciless fanatics: her flesh was scraped from her bones with sharp oyster-shells and her quivering limbs were delivered to the flames. (Gibbon, *Decline & Fall*, Ch. 47)

Russell (*HWP*, p. 368) adds that "Alexandria was no longer troubled by philosophers."[2]

Destruction of the Museum

The final blow to Alexandrian scholarship in the west came in 640, when the Moslems captured Alexandria and destroyed the famous Museum (essentially a library and research institute); it is said that the public baths were heated for six months by burning manuscripts. By then, however, most intellectuals had fled to other lands, bringing with them copies of important manuscripts.

The Closing of the Academy

The persecution of Pagan learning continued throughout the empire, and in 529 Justinian ordered the closure of the Pagan schools, including the Academy, founded by Plato:

> We forbid anyone stricken with the madness of the impure Hellenes [Pagans] to teach, so as to prevent them, under the guise of teaching those who by misfortune happen to attend their classes, from in fact corrupting the souls of those they pretend to educate. (Cod. Just. 1.11, law 10, quoted in Chuvin, *CLP*, p. 133)

The Academy's endowment was confiscated in 531, and seven of its faculty fled to Mesopotamia, where they had been invited to join the court of Chosroes, emperor of the Sassanians, who were Zoroastrian rather than Christian. Some of these scholars founded colleges in Persia, as did Nestorian Christians, who had been declared heretical and so fled.

The Rise of Hindu-Arabian Mathematics

Arabian mathematics acquires some of its distinctive characteristics from having developed in the midst of several ancient civilizations. Of course there was a long indigenous mathematical tradition in Mesopotamia and Babylonia, but the Arabs were also subject to influences from China, India, Egypt and, beginning in the third century CE, the Alexandrian Greeks. After an initial spate of religious fanaticism and intolerance, the Moslems became more tolerant of the learning of the "infidels," and soon the Pagan and heretic

[2]Damascius (6th cent.) thought that Cyril was threatened by Hypatia's popularity with students, but another explanation is that he saw this as a way of having revenge on her pupil Orestes, prefect of Alexandria, who was trying to suppress Cyril's anti-Semitic rabble-rousing.

refugee-scholars began translating works of Greek science, including Aristotle and Euclid, into Syriac, Pahlavi (an Iranian script) and Arabic.

Blurring the Number/Magnitude Distinction

The Arabs were more interested in using mathematics for practical purposes, in commerce and astronomy, than with its theory, and so they emphasized arithmetic more than geometry. Arabian applications of mathematics were aided by their adoption of the Hindu number system (the predecessor of our own decimal numbers). The Hindus had developed a positional number system in the sixth century CE, and it was brought to Arabia c. 825 by al-Khowārizmī (about whom more later). Perhaps because of their focus on calculation rather than proof — or perhaps because they were less under the domination of *logos* — the Arabs were not as worried about irrational numbers as were the Greeks. Both Omar Khayyām (c. 1100) and Nasīr-Eddin (1201–1274) took all ratios — whether commensurable or incommensurable — to be numbers. Also, the Hindu mathematician Bhāskara II (born, 1114) derived rules for arithmetic on surds by assuming that they follow the same rules as integers. He had no proof that these rules were correct, but this is just one example, among many, of mathematics progressing not in spite of, but because of an ignorance of the technical problems.

Facility with Irrationals

As a result of these conditions, Arab mathematicians were generally comfortable with irrational numbers, and so did not draw as sharp a distinction between arithmetic and geometry as did the Greeks. When, beginning about 1100, Arabian mathematics was brought to Europe, its acceptance of irrationals helped to break down the boundary between arithmetic and geometry, which Europe had inherited from the Greeks. By blurring the boundary, Arabian mathematics opened the way for the development of analytic geometry.

By the sixteenth century, irrational numbers were freely used in European mathematics and their properties were becoming familiar. For example, Michael Stifel (c. 1486–1567) knew that irrational numbers have an infinite decimal expansion, and Simon Stevin (1548–1620) used rational numbers to approximate them. Despite a lack of theoretical justification, mathematicians discovered that they could work with irrational numbers as well as with integers, so the arguments for banning them from arithmetic lost some of their force.

Equal Temperament

The growing acceptance of irrational numbers in mathematics was paralleled by their incorporation in musical theory.[3] There were problems with

[3]Sources include Wiener (*DHI*, Vol. 3, pp. 260b–264a), Ronan (*Science*, pp. 314–315) and Cohen & Drabkin (*SBGS*, pp. 286–307).

the Pythagorean tuning system (Section 2.2.1) when it was extended, since twelve fifths differed by about a quarter of a semitone from seven octaves:

$$\left(\frac{3}{2}\right)^{12} : \left(\frac{2}{1}\right)^{7} \;::\; 531\,441 : 524\,288.$$

To compensate for this inconvenient misalignment, musicians, tuning their instruments by ear, “tempered” their fourths and fifths by adjusting their pitches slightly away from the ideal ratios given by Pythagoras. This had the disadvantage of making the thirds and sixths more dissonant, but they were little used in ancient music. By the fourth century BCE Aristoxenus of Tarentum, a student of Aristotle’s, had argued that ratios are irrelevant to music, and he suggested a subjective division of the octave; Ptolemy (c. 100–170 CE) proposed a third tuning (based on oblong numbers) that was considered a compromise between the extreme rationalism of Pythagoras and the extreme empiricism of Aristoxenus.

In the sixteenth century the gap between music practice and music theory widened as composers began to make greater use of thirds and sixths. Support for Pythagorean tuning was further weakened when Vicenzo Galilei (Galileo’s father) discovered in 1589–1590 that although the Pythagorean ratios apply to string and pipe lengths, other ratios apply to other physical magnitudes affecting pitch. Further, Giovanni Battista Benedetti (1588–1637) explained consonance in terms of the number of coincident peaks between two tones in a given time interval, which implied consonance versus dissonance was a matter of degree rather than a yes-or-no question. His theory that there is a *continuum* of intervals gained the support of Isaac Beeckman (1588–1637) and Marin Mersenne (1588–1648) in the seventeenth century.

All these developments loosened the hold of the Pythagorean system and paved the way for the publication around 1620 by Stevin of the *equal temperament* system. He may have got the idea from a similar system published in 1585 by a Chinese prince, Chu Tsai-Yu (Zhu Zai-You), or he may have rediscovered it independently. Equal temperament divides the octave into twelve equal semitones, each with a pitch ratio of $\sqrt[12]{2}$. This even division was much more convenient for the more distant modulations, semitone melodic movements and more dissonant chords that were becoming popular at that time. Bach’s *Well-Tempered Clavier* is intended to illustrate the flexibility of the new system. Clearly a system based on a ratio of $\sqrt[12]{2}$ could not have been devised, had mathematicians not become accustomed to irrational numbers.

5.1.2 Analytic Geometry

> I have resolved to quit only abstract geometry, that is to say, the consideration of questions which serve only to exercise the mind, and this, in order to study another kind of geometry, which has for its object the explanation of the phenomena of nature.
>
> — Descartes

> Whenever in a final equation two unknown quantities are found, we have a locus, the extremity of one of these describing a line, straight or curved.
>
> — Fermat

Precursors to Analytic Geometry

René Descartes was not the first to graph curves in a rectilinear coordinate system; this had been done in cartography (latitude and longitude) and astronomy for a long time.[4] Also Omar Khayyām and other Islamic mathematicians had recognized a relationship between algebra and geometry. Later (1346) Giovanni di Casali explored the graphical representation of equations of motion, a method that was refined by Nicole Oresme (c. 1350), and later used by Galileo (1638). However, it was Descartes who first showed the essential identity of algebraic expressions and geometric curves, although Fermat had the idea about the same time, but didn't publish it (Clagett, *SMMA*, pp. 331–333).[5] This major advance in mathematics was published in Descartes' *La Géométrie*, which was an *appendix* (along with *La Dioptrique* and *Les Météores*) to his *Discours de la Méthode* (1637), apparently to illustrate the application of his Method. At first Descartes' geometry was

[4] The earliest known graph is a plot of the change of latitudes of the planets relative to their longitudes, which appears in an eleventh century manuscript (Munich 14436, reproduced in Boyer, IAG, and in Crombie, *MEMS*, Vol. 2, plate 2). The principal secondary source for this section is Kline (*MT*, Chs. 13, 15).

[5] The idea of representing time as a line goes back at least to Aristotle (*Physics* Bks. 4, 6, 7). Decartes first described the idea of coordinate geometry in a 1619 letter to Isaac Beeckman, although his *Geometry* wasn't published till 1637. Pierre de Fermat (1601–1665) had the idea of coordinate geometry 1629, but his work was not published till 1679, after his death, although it circulated in manuscript in the meantime. A bitter dispute resulted from claims of priority in what was, typically, a parallel development of ideas that were "in the air," a foreshadowing of the Newton/Leibniz dispute over the calculus.

criticized by Fermat, Roberval and Pascal, but eventually it was accepted and became an important tool in science.[6]

Use of Linear Magnitudes

Descartes' *Géométrie* begins:[7]

> Any problem in geometry can easily be reduced to such terms that a knowledge of the lengths of certain straight lines is sufficient for its construction.

He further claims that in reducing geometry to lengths of lines, only four or five operations are required, and these operations are analogous to the arithmetic operations ($+$, $-$, $\times$, $\div$, $\sqrt{}$). Although we think of his accomplishment, analytic geometry, as the reduction of geometric curves to equations on real numbers, he thought of it differently, because real numbers, per se, were not yet invented. Descartes developed analytic geometry in the context of magnitudes.

Problems with Combining Magnitudes

We saw in Section 2.6.3 how the apparently irreconcilable rift between the rational and the irrational led to a distinction between number (*arithmos*) and magnitude (*megethos*), and to an essentially independent axiomatization of arithmetic, the science of discrete quantity, and geometry, the science of continuous quantity.[8] Although mathematicians were becoming more comfortable with irrationals, this basic framework was still in place when Descartes developed his geometry.

Combination of Magnitudes

The Eudoxan/Euclidean theory of magnitudes permitted the combination of magnitudes — whether by addition, subtraction, multiplication or division — only if they were of the same kind. For example, lengths could be added to lengths, or times to times, but a length could not be divided by a time. This restriction impeded the development of mechanics for 2000 years; indeed, the modern notion of the *dimensions* of a quantity was formulated as recently as 1863, by Maxwell and Jenken (Bochner, *RMRS*, pp. 209–212; Dijksterhuis, *MWP*, p. 192).

The Problem with Multiplication

---

[6]Descartes' ideas on scientific method were very influential among seventeenth century intellectuals, although the Church could not accept his view that reason could settle theological issues. Shortly after his death his books were placed on the *Index of Prohibited Books*. (Indeed, fear of Church prosecution had convinced him not to publish one of his earliest works, *Le Monde*, which presents a cosmological system.)

[7]The primary source for this section is Descartes' *La Géométrie* (Descartes, *Geom*); the first book is reprinted and translated in Newman (*WM*, Vol. 1, pp. 242–253). A concise history of analytic geometry is Boyer (IAG).

[8]Aristotle, *Cat.* 4b20–5a39, *Post. An.* 75a35–76a16.

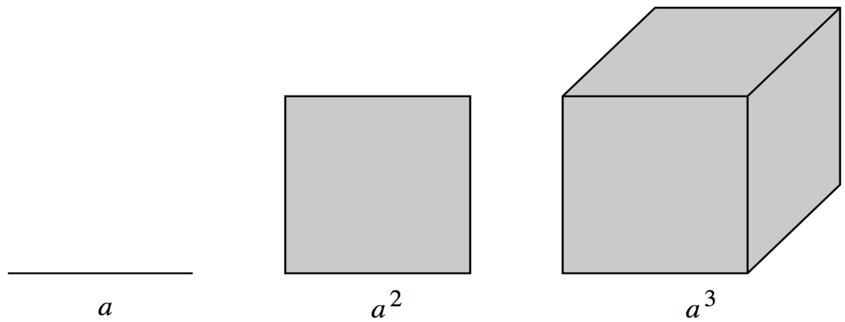


Figure 5.1: Squares and Cubes of Lines. In the Eudoxan/Euclidean theory of magnitudes, the square a^2 of a line a was taken to be a square in fact, and its cube a^3 to be a cube in fact.

There was an additional complication: the multiplication of lines was taken to produce surfaces, the multiplication of lines by surfaces to produce volumes, and the multiplication of surfaces to be meaningless; the square of a line was literally a square, and its cube literally a cube (Fig. 5.1). This limited the Greeks' ability to deal with polynomials of more than the third degree.

Weakening of Distinction

Partly under the influence of Arabic mathematics, the distinction between numbers and magnitudes was beginning to weaken. For example, the brilliant Niccolò "Tartaglia" (The Stammerer) Fontana (c. 1500–1557) complained that some mathematicians were confusing *multiplicare*, the multiplication of numbers, with *ducere*, the multiplication of magnitudes. On the other hand, François Viète (1540–1603)[9] argued that algebra was better then geometry, because it was not limited to equations of the third degree or lower. In fact, Decartes claimed that he began where Viète had stopped.

Descartes' Multiplication

Descartes' goal was to reduce geometry to a systematic method, and to do this he had to devise more general methods of combining magnitudes, but he was still in the Greek framework: the only *numbers* were positive integers and ratios of positive integers. Figure 5.2 shows Descartes' method for multiplying (linear) magnitudes. Suppose we want to multiply the length BC by the length BD and get a *length* as the result. To accomplish this it is necessary to pick some length u to be the unit; although Descartes says it can usually be chosen arbitrarily, the need for it is significant and will be

[9] Also, Franciscus Vieta.

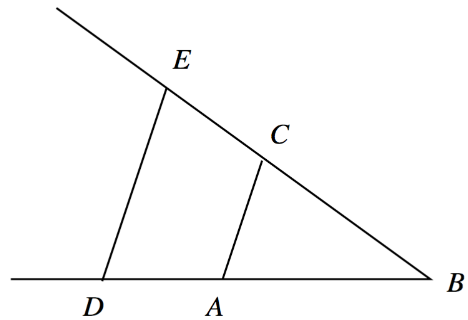


Figure 5.2: Descartes' Method for Multiplying Magnitudes. The goal is to construct a line whose length is the product of BC and BD. A "unit" AB is chosen arbitrarily and DE is drawn parallel to AC. Then BE is the required product, since it satisfies the proportion $BE/BC = BD/1$. Division is accomplished analogously.

addressed later. Once a unit has been chosen, the required product can be expressed as a length p satisfying the proportion $p : BC \;::\; BD : u$; in modern notation, $p/BC = BD/1$, so $p = BC \times BD$. The required length is easily constructed: Set out the unit BA along BD and connect A and C. Then draw DE parallel to A; by similar triangles BE satisfies the proportion $BE : BC \;::\; BD : BA$, so it is the required product.

Division and Square Root

To divide BE by BD is just as easy: join E and D and construct AC parallel to it; BC is the desired quotient. Decartes discusses several other operations — extracting the square root is illustrated in Fig. 5.3 — including the solution of quadratic equations and "Descartes' Rule of Signs" (actually discovered by Thomas Harriot, 1560–1621):

> But I will not stop to explain this in more detail, because I should deprive you of the pleasure of mastering it yourself, as well as of the advantage of training your mind by working over it, which is in my opinion the principal benefit to be derived from this science.
>
> — Descartes (*La Géométrie*, Bk. 1)

No Unit in Greek Multiplication

It is worthwhile to consider the need for a unit. Recall that a magnitude is a quantity, for example, the length of this line:

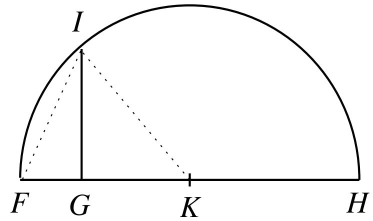


Figure 5.3: Descartes' Method for Extracting Square Roots. The goal is to extract the square root of GH. Extend this line to F so that FG is equal to unity. Bisect FH at K and draw the semicircle FIH. The length of the perpendicular line GI is the square root of GH. To see this, let $x = GH$, $u = FG$, $h = KI = KF$ and $r = GI$. By the Pythagorean theorem, $r^2 = h^2 - \overline{GK}^2 = [(x+u)/2]^2 - [(x+u)/2 - u]^2 = xu$, so $r = \sqrt{x}$.

It is not a *number* of inches, millimeters or any other units. Although magnitudes of the same kind can be added or subtracted and yield magnitudes of the same kind, multiplication and division are different. Multiplication of magnitudes produces magnitudes of higher dimension, so for example a length times a length produces an area (a length squared). Since multiplication does not produce a magnitude of the same kind as its inputs, it doesn't make sense to ask if it has a unit, since by definition there is no u such that $ux = x$ (there is *no* area equal to the line displayed above). Further, ratios of magnitudes did not produce a magnitude at all; the Greeks had no concept of a *pure* (i.e. dimensionless) magnitude. Ratios of compatible magnitudes were only allowed in a proportion, where they are compared to each other. In a proportion, ratios of magnitudes of different kinds can even be compared, as, for example, when they said that the areas of circles are in the same ratio as the lengths of their radii. But ratios were considered neither numbers nor magnitudes, so they could not be combined arithmetically.[10]

Unit is Arbitrary

When we come to define a multiplication operation on magnitudes that yields magnitudes of the same kind, it is apparent that we need a unit, for one of the most basic properties of multiplication is that it has an identity, that is, a unit u such that $ux = xu = x$. It is also apparent that this unit is arbitrary,

[10]It is interesting that allowing ratios as pure magnitudes forces the existence of a unit, since for any magnitude x, $x/x = 1$.

for if we consider lengths, for example, there is no reason to prefer one length over another for the unit. In other words, the only *distinguished* length is the zero length, but that cannot be the unit, because it is the *annihilator* for multiplication ($0x = x0 = 0$). So we have the curious situation that the ability to multiply magnitudes presumes a unit, a distinguished unit of measure for magnitudes of that kind.

The Unit Generates the Integers

Once we have specified a unit, our system of magnitudes also contains the integers, which can be generated by adding the unit to itself repeatedly, $1 = u$, $2 = u + u$, $3 = u + u + u$, etc. Thus, by Descartes' device we have generated the integers from the continuum, we have *geometrized arithmetic*! This is not however the route mathematics followed, so strong was the bias toward the discrete. Nevertheless, it is an informal demonstration, which, combined with the later arithmetization of geometry, shows that either the continuous or the discrete may be taken as basic, and the other reduced to it.

Cartesian Coordinates

Descartes observes in a letter to Princess Elizabeth that he takes care, "as far as possible, to use as lines of reference parallel lines or lines at right angles," that is, to use Cartesian coordinates (Newman, *WM*, Vol. 1, p. 249). Thus, not only did he reduce geometrical problems to simple operations on lines, Descartes also showed how all the lines could be in two perpendicular directions, that is, the x and y axes. This is the invention of Cartesian coordinates (although we've seen that latitude and longitude was a familiar idea).

Use of a Systematic Method

Note that Descartes did not accomplish the arithmetization of geometry, the reduction of the continuous to the discrete. He reduced geometry to operations on magnitudes, which are themselves geometric (continuous) objects. So, in effect he reduced geometry to a very simple subset of geometry. This was important for two reasons. First, it showed how complex geometrical problems could be solved by a systematic process involving only a few arithmetic-like operations. (He goes so far as to say, as we'll see shortly, that he will use arithmetical terms for the operations on lines.) Second, it showed how geometry could be arithmetized *if only* continuous line segments (i.e., real numbers) could be reduced to integers. We'll see how this was accomplished in Section 5.3.

5.1.3 Algebra and the Number System

Use of Letters for Quantities

Aristotle seems to have originated the use of letters to indicate variable quantities; I have already mentioned their use in logic (p. 52), but he also used them in physics.[11] Algebra was advanced further in Europe by the importation of Arabian techniques, including the Hindu-Arabic number system. The word 'algebra' derives from Arabic *al-jabr*, which means *restoration* and refers to transposition of terms across an equation. One of the main sources was the book *Al-jabr wa'l-muqābaba* (*Restoration and Equation*), by Abū Ja'far Muhammad ibn Mūsā al-Khwārizmī (c. 800–850), who brought Hindu numbers to Arabia and from whose name we also get the word 'algorithm'.

Use of Letters for Constants

Although algebraic notation developed through the sixteenth century, letters were still used only for *variable* quantities, until Viète began to use vowels for unknowns and consonants for constants. This may have given the idea to Descartes (who rarely acknowledged sources), since he adopted the convention of using letters near the end of the alphabet (x, y, z) for unknowns and letters near the beginning (a, b, c) for constants.

The use of letters for constants is an important contribution, since it encourages one to look at equations in general. That is, specific equations, such as $15x^2 + 6x - 2 = 0$ or $3x^2 - 37x + 9 = 0$, incline one to treat each as a separate problem. But when we write the equation $ax^2 + bx + c = 0$, then we are more likely to develop methods that are independent of the specific constants a, b and c. Instead of seeing isolated problems, we see them as members of a class.

Negative Magnitudes

Less obviously, the introduction of algebraic notation encouraged the extension of the number system. Descartes uses the terminology of arithmetic (addition, subtraction, multiplication, division, square root) to denote operations on magnitudes, and he uses algebraic expressions such as $\sqrt{a^2 + b^2}$ to describe magnitudes, so he had already blurred the distinction between arithmetic and geometry, for in many cases the same formula could describe either numbers or magnitudes. This would encourage their unification.

Confusion about Negatives

It is perhaps surprising that in Descartes' time negative numbers were not generally accepted, but it is more understandable if we think of them in Greek terms. How could you have a fewer than no pebbles in a Pythagorean figure? Or a negative length or area? Even zero was something of a second-class citizen, both as a number and a magnitude; they were understandably

[11] Principle sources for this section are Kline (*MT*, Chs. 9, 13), Ronan (*Science*, pp. 149, 192, 321–323) and Newman (*WM*, Vol. 1, pp. 118–120).

uncomfortable thinking of a figure with no pebbles as a figure, or a line with no length as a line.[12] As early as the seventh century, Brahmagupta had used negative numbers to represent debts, and in the twelfth century the nineteen year old ibn Yahya al-Samaw'al used negative numbers in the solution of equations, but Omar Khayyām and Nasīr-Eddin both rejected them. In Europe, Nicolas Chuquet allowed them as roots in his unpublished 1484 book, but they were accepted in differing ways by sixteenth and seventeenth century mathematicians. For example, Viète and Pascal reject negatives outright, but Cardano (1501–1576), Stevin and Albert Girard (1595–1663) allow them as roots of equations, but don't consider them numbers; in contrast, Harriot took them to be numbers, but didn't permit them as roots. John Wallis (1616–1703), the century's best British mathematician after Newton, accepted negative numbers, but thought they were greater than infinity! Descartes did not allow them, and in his *Geometry* he says that a quadratic with two negative roots has *no* roots. In the seventeenth century the negative roots of equations were often called *false* or *fictitious* roots (as opposed to affirmative roots), and even in the nineteenth century some scientists (e.g. Carnot, DeMorgan) still claimed that negative numbers were a sign of errors, inconsistencies or absurdities.

Variables Hide Negatives

Now the interesting thing about algebraic notation is that when you write $ax^2 + bx + c = 0$ you don't know (or can choose to ignore) whether the numbers a, b and c are positive, negative or zero. Thus the use of letters for constants helped mathematicians to see that negative numbers worked just as well as positive, and in fact that many procedures were simplified if letters were allowed to stand for negative numbers. Mathematicians are actually rather pragmatic, and if something works consistently enough it will eventually be accepted; it has happened repeatedly (e.g., negative numbers, imaginary numbers, infinitesmal calculus, infinite series, Fourier analysis, vectors,

[12] Zero was introduced to Europe in 1202 by Leonardo Fibonacci of Pisa in *Liber abaci*, the first algebra book in Latin, although it wasn't widely accepted until about 1500. Although the Alexandrian Greeks had a symbol for zero, it wasn't accepted as a full-fledged number, which occurred among Hindu mathematicians in the seventh century CE. They may have got the idea from Cambodia or Sumatra, where it was more acceptable as the "void" or "emptiness" of Taoism and other Eastern philosophies (Ronan, *Science*, pp. 149, 192, 321–323). It is worth remembering that in ancient Greek times, the numbers didn't include 1, which was neither odd nor even (Maziarz & Greenwood, *GMP*, p. 18). (In everyday speech 'number' still usually means *more than one*: "I've read a number of books." "How many?" "One.") We see though history a progressive extension of the class of numbers.

generalized functions). Successful practice regularly overrules metaphysical qualms.

Imaginary Numbers

Imaginary numbers came to be accepted in a similar way to negatives. It is interesting that although negative numbers are now much more familiar to non-mathematicians than are imaginary numbers, they entered the number system about the same time.[13] There was a practice in sixteenth century Italy of mathematicians challenging one another to public debates with financial rewards. Therefore they were motivated to keep their methods secret, even at the risk of not receiving recognition for a discovery.[14] Thus Tartaglia passed — under an oath of secrecy — certain methods for solving cubic equations to Girolamo Cardano (1501–1576), who nevertheless published them in his *Ars magna* (1545).[15] In this book he works through the solution of a quadratic equation with complex roots. He observes that "putting aside the mental torments" (*dimissis incruciationibus*) we can carry out the operations to solve the equation, but that the resulting solution is "as subtile as it is useless" (*est subtile, ut sit inutile*).

Rafael Bombelli (1572) and Viète (1591) discussed the case in which *imaginary* (also called *impossible*) numbers appear in the solution of cubic equations that have real roots. The paradox is that even though the method made use of "imaginary" quantities, the solutions were real and could be checked; they were found to be correct. It would have been easier to reject imaginary

[13] The Hindu mathematician Mahāvīra (fl. 850 CE) accepted $\sqrt{-1}$ as the root of an equation, but his idea doesn't seem to have had any influence.

[14] The idea of keeping discoveries secret, which seems so alien to contemporary academic research, was much more common at that time. Many scholars kept their discoveries secret, or wrote them in code in an attempt to have both secrecy and a proof of priority. Another reason for this practice may be the tradition of secrecy surrounding "occult knowledge"; recall that from antiquity to the end of the seventeenth century, the border between magic and science was indistinct at best, and many scholars practiced both (e.g., Roger Bacon, Bruno, Tycho, Kepler, Leibniz, Newton). See also Section 5.2.

[15] Also Hieronymus Cardanus or Jerome Cardan. Ronan (*Science*, pp. 322–323) adds this about Cardano: "Imprisoned by the Inquisition in 1570, mainly for the horoscopes he had cast, he recanted and succeeded in obtaining a lifelong annuity from the Pope." Newman (*WM*, Vol. 1, p. 119) calls him "a turbulent man of genius, very unscrupulous, very indiscreet, but of commanding mathematical ability," and says "he combined piracy with a measure of honest toil." Among other accomplishments, he wrote the first probability text (*Liber de ludo aleae*). To his credit he *did* acknowledge that Tartaglia gave him the method (in obscure verses!), but also stated that it was identical to the method previously devised by del Ferro (c. 1500); in addition Tartaglia did not provide a proof, which was done by Cardano (Struik, *SBM*, pp. 62–67).

numbers if they didn't given such useful results! Mathematicians were discovering more and more cases in which, by "dismissing the mental torments" and manipulating apparently meaningless symbols, they could achieve useful results. Leibniz said (*Acta erud.*, 1702; Kline, *MT*, p. 254):

> The Divine Spirit found a sublime outlet in that wonder of analysis, that portent of the ideal world, that amphibian being being and not-being, which we call the imaginary root of negative unity.

Yet by the end of the eighteenth century, negative and imaginary numbers were generally accepted. As each mooring line to reality was cut, mathematics was able to fly higher and farther.

> The motto which I should adopt against a course which seems to me calculated to stop the progress of discovery would be contained in a word and a symbol — remember $\sqrt{-1}$.
>
> — DeMorgan, *Differential and Integral Calculus*

Formalization and the Extension of the Number System

We will see other examples where notation has driven the extension of the number system or led to the creation of other new mathematics, or, as we might say, cases where *syntax generates semantics*. The principal example is the infinitesmal calculus, discussed in Section 5.3, but Cantor's set theory is also important (Section 6.1.3). All these focused attention on *formality* in mathematics, which eventually gave birth to the theory of computation.

5.1.4 The Importance of Informality

The Changing Roles of Algebra and Geometry

On one hand, Descartes showed how geometry could be reduced to arithmetic-like manipulation of linear magnitudes, which was more systematic than Euclidean geometry and not restricted to three dimensions. On the other, the practical but nonrigorous algebra of the Arabs showed that irrational numbers could be manipulated as reliably as the integers. In addition to blurring the distinction between arithmetic and geometry, the greater facility of algebraic techniques caused more people to adopt its methods. The trouble was that geometry provided the only rigorous way to manipulate irrationals. As more and more use was made of algebra, many seventeenth century mathematicians became concerned about algebra's lack of a rigorous foundation. We've seen how Descartes still resorted to Euclidean techniques of proof, but he nevertheless saw algebra as a powerful tool of thought that is, in a sense, prior to mathematics. Blaise Pascal (1623–1662) and Isaac Barrow (1630–1677) objected to the use of algebraic techniques in geometry, as did Newton: "Equations are expressions of arithmetical computation and properly have no place in geometry" (*Arith. univ.*, 1707, p. 282; Kline, *MT*, p. 318).

Complementarity

Nevertheless, most mathematicians were unconcerned with logical quibbles, and the progress of algebra continued, to the benefit of mathematics and science:

> As long as algebra and geometry traveled separate paths their advance was slow and their applications limited. But when these two sciences joined company, they drew from each other fresh vitality, and thenceforward marched on at a rapid pace towards perfection. (Lagrange, *Leçons élémentaires sur les mathématiques*; Kline, *MT*, p. 322)

This synergy results from their complementary viewpoints, since the unification of algebra and geometry permitted mathematicians and scientists to move gracefully between the two, adopting the algebraic view when it was more productive (e.g., for manipulation), and the geometric when it was better (e.g. for visualization). Progress would have been blocked by excessive concern for the unresolved irrationalities. As Kline (*MT*, p. 282) observes,

> And it is fortunate that the mathematicians were so credulous and even naive, rather than logically scrupulous. For free creation must precede formalization and logical foundations, and the greatest period of mathematical creativity was already under way.

Nevertheless, Zeno's objections could not be ignored indefinitely; the continuum had not been rationalized.

5.2 Magic and the New Science

5.2.1 Pythagorean Neoplatonism

> God in creating the universe and regulating the order of the cosmos had in view the five regular bodies of geometry as known since the days of Pythagoras and Plato.

— Kepler, 1595

Neoplatonic-Pythagorean Revival

We are on the threshold now of the scientific revolution, and it will be worthwhile to consider three catalysts that triggered its precipitous growth. The first of these was the growing interest in the fourteenth century in Renaissance Neoplatonism, which included a significant amount of mystical Pythagoreanism. Since Aristotle and Plato have often been seen as opposing poles in philosophy, the reaction against Aristotle, as represented by the Schoolmen, led to the promotion of Plato, as represented by the Renaissance Neoplatonists.

History of Neoplatonism

Neoplatonism has existed in a continuous tradition from the time of Plato. His immediate successors as head of the Academy (Speucippus, c. 407–339 BCE, and Xenocrates, died c. 314 BCE) emphasized the mathematical aspects of Plato's philosophy, and in this sense they returned to early Pythagoreanism. After about two centuries of a more skeptical approach, the Academy returned to more metaphysical interests, and created an eclectic blend of Platonic, Aristotelian, Neopythagorean and Stoic ideas that constituted the basis of Neoplatonism. Neoplatonic thought was absorbed by a number of the Church Fathers, especially St. Augustine (354–430 CE), and permeated their writings, which may have helped it survive the suppression of Pagan learning. Neoplatonism reached its characteristic form in Plotinus (c. 205–270 CE), who — perhaps unconsciously — also added some gnostic elements, and in Iamblichus (c. 250–c. 325 CE) who extended its metaphysical system and emphasized the role of ritual magic.[16] Two other Neoplatonists were also very influential on medieval thought: Proclus (c. 410–485), a Pagan philosopher and head of the Academy, and Dionysius (c. 500), mistakenly identified with the Areopagite, but more influential because of that mistake.

From the mid-fifth to the mid-sixth centuries there were two main Neoplatonic schools, the Academy in Athens and the Alexandrian School. Neoplatonic speculation at the Academy ended when Justinian closed the Pagan schools in 529. The Alexandrian school survived by Christianizing itself, abandoning ritual magic, and concentrating on textual criticism rather than philosophy. Thus it became the main vehicle by which Neoplatonism entered medieval thought.

[16] Jonas (Gn) is a good introduction to gnosticism.

The enshrinement of Neoplatonism in the thinking of a number of theologians, including Augustine and Boethius (c. 480–524), as well as in other widely read authors such as Macrobius (c. 400), helped keep it alive for some 1000 years (250–1250). Although Neoplatonism suffered an eclipse with the rise of Scholasticism, which found Aristotelian logic more suited to its needs, it flowered again in the hands of the Renaissance Humanists beginning with Petrarch (1304–1374).

Renaissance of Neoplatonism

Neoplatonism had an important influence on many of the founders of the scientific revolution, including Copernicus (1473–1543), Kepler (1571–1630), Galileo (1564–1642) and Descartes (1596–1650).[17] This is because Pythagorean Neoplatonism — in contrast to Aristotle's science, which was largely qualitative — stressed the importance of mathematics as the path to the secrets of the universe. Furthermore, the Neoplatonic view that the forms, and thus the mathematical laws of the universe, were ideas in the mind of God made the pursuit of science seem more like religious practice, and so more acceptable in those often intolerant times.

Pythagorean Neoplatonism continued to exert its mathematicizing influence on science until its decline at the end of the seventeenth century, but by then mathematical science was well launched as an independent discipline, and indeed was already at pains to separate itself from "Neoplatonic pseudoscience."

5.2.2 Hermeticism

> All these things, as many as there are, have been revealed to you, Thrice Great One. Think through all things yourself in the same way,

[17]Kepler considered his most important contribution to be his discovery that the five Platonic solids fit between the orbits of the six planets. His *Musical Theory of the World* (*Harmonice mundi*, 1619) contains an astonishing discussion of the harmony of the spheres, which he takes literally, not metaphorically. In true Pythagorean fashion he works out the exact scales, and in Ch. 8 of Bk. 5 asks, "In the Celestial Harmonies, which planet sings soprano, which alto, which tenor, and which bass?" Answer? Saturn and Jupiter sing bass, Mars tenor, Earth and Venus alto, and Mercury soprano! His belief in the harmony of the universe led him to develop a system of astrology, which he considered "an adjunct and an ally to astronomy," and he frequently cast horoscopes; his almost legendary reputation in Graz, where he was a mathematics professor, was established by some accurate predictions he made of the weather and peasant uprisings. Religious persecution forced Kepler to flee Graz and to join Tycho at Prague in 1600. See Clulee (*JDNP*, p. 73), French (*JD*, pp. 1–2), Kepler (*GB*, biog. note, p. 841), Ronan (*Science*, pp. 337–338) and Wightman (*SiRS*, pp. 34, 146–148).

and you won't go wrong.

— *Corpus hermeticum*, 11.2

Or let my lamp, at midnight hour,
Be seen in some high lonely tower,
Where I may oft outwatch the Bear,
With thrice-great Hermes, or unsphere
The spirit of Plato, to unfold
What worlds or what vast regions hold
The immortal mind that hath forsook
Her mansion in this fleshy nook;
And of those Daemons that are found
In fire, air, flood, or underground,
Whose power hath a true consent
With planet or with element.

— Milton, *Il Penseroso*

Where is Hermes Trismegistus,
Who their secrets held?

— Longfellow, *Hermes Trismegistus*

The One remains, the many change and pass;

— Shelley, *Adonais* 52

Hermetic Influences

The second major impetus to the scientific revolution is perhaps more surprising than Neoplatonism, but its importance has been well defended by Frances A. Yates and her students and colleagues at the Warburg Institute (University of London).[18] It is *Hermeticism*, or the philosophy deriving from the writings attributed to "Hermes Trismegistus." When translations of these first became available to Europe in the fifteenth century, they were

[18]See, for example, Yates (*AoM*, *GB&HT*, *OPEA*, *RE*) and French (*JD*). Another source for this section is Clulee (*JDNP*). Yates (Herm) is a brief overview of Hermeticism.

believed to be a very great antiquity — they were dated to the time of Moses — and were considered a kind of revelation from the Egyptian god "Thrice Great" (i.e. Very Great) Thoth, whom the Greeks equated with Hermes.[19] Seventeenth-century scholarship showed that they were no older than the second and third centuries CE, and after that they lost some of their influence, but in the meantime they had had an important impact on the new science. Also surprising is the fact that the Hermetic writings had received the endorsement of the Church Fathers, especially Lactantius (born, 250 CE) and Augustine, apparently because they saw Hermes Trismegistus as the foremost "Gentile prophet" who had foretold the coming of Christianity. Perhaps this preserved the *Hermetica* from suppression; in any case they had an enormous influence on the Renaissance.

Hermetic Influence in Arts

In addition to science, which is our concern here, Hermetic influence is apparent in art (e.g., Botticelli's "Primavera," Dürer's "Melancholia," Pinturicchio's frescoes in the Borgia Apartments in the Vatican), poetry (Chapman's *Shadow of Night*, Spenser's *Faeire Queene*, Sidney's *Defense of Poesie*), drama (Shakespeare's late plays, the Globe theater) and philosophy (Giordano Bruno).[20] Roger Bacon (c. 1214–1292) — the father of experimental science — called Hermes Trismegistus "the father of philosophers" (Thorndyke, *HMES*, Vol. 2, p. 219).[21]

Origin of Hermetica

Current opinion is that a number of Greek writers composed the *Hermetica* between 100 and 300 CE, and that these writings reflected the popular philosophy of the time, which combined Neoplatonism and Stoicism, with a small admixture of Jewish and Persian ideas; it is built on a framework of Hellenistic astrology. One attitude that permeates this philosophy is a shift from the rational dialectic of the Greeks, which seemed incapable of settling the important philosophical issues of the day, to *gnosis*: an intuitive apprehension of the divine and natural worlds and of humanity's place in them.

[19] The writings themselves were supposed to have been authored by an "ancient theologian" (*priscus theologus*) whom they called "Hermes Trismegistos."

[20] See especially Yates (*OPEA*); she says that *The Tempest* "presented in Prospero, magus and conjurer, Shakespeare's last word on occult philosophy." Feinstein (*DHI*, Vol. 2, pp. 433a–434a) adds to the list: Milton, Sterne, Blake, Longfellow, Goethe, Wordsworth, Nietzsche, G. M. Hopkins and D. H. Lawrence.

[21] After joining the Franciscan Order late in life, Bacon found himself in a bind, for he had been secretly commissioned by the pope to prepare a work on his "universal science," but the Franciscans had placed him under supervision and forbidden him to write. He completed the work anyway, but after the accession of a new pope he was imprisoned for "certain suspected novelties" (Jones, *HWP*, Vol. II, p. 290).

Yates (*GB&HT*, p. 5) calls Hermeticism "a religion, a cult without temples or liturgy, followed in the mind alone, a religious philosophy or philosophical religion containing a gnosis."

Supposed Antiquity of Hermetica

The *Hermetica*'s reputation for great antiquity must have been established early, since Lactantius and Augustine both take it for granted, and it was the unquestioned acceptance of their opinion that lent credibility to the *Hermetica*, for the Renaissance valued nothing more than ancient wisdom, believing it to be purer and holier, that "the earliest thinkers walked more closely with the gods than the busy rationalists, their successors" (Yates, *GB&HT*, p. 5). Because of their assumed antiquity, historical references in the Hermetica were seen instead as *fulfilled prophecy*, which increased the credibility of the writings even more. As Yates (*GB&HT*, p. 6) says, "This huge historical error was to have amazing results."

Renaissance of Hermeticism

The *Hermetica* were lost to Western Europe until a manuscript was brought to Cosimo de' Medici around 1460. (It may have come with refugees from Byzantium after the fall of Constantinople to the Turks in 1453.) Marsilio Ficino (1433–1499) was ordered by Cosimo to drop his project of translating Plato and to work instead on the *Hermetica*, so that the 70 year-old Cosimo might read them before he died. After Ficino completed the translation, he devoted himself to the further development of Hermeticism, combining it with elements of Neoplatonism and Orphism. In true Renaissance style, he gave his ideas a remarkable pedigree, saying that the knowledge had been passed down from Zoroaster and Hermes Trismegistus to Orpheus, from Orpheus to Aglaophemus, and from him to Pythagoras, whence they came to Plato (perhaps through Philolaus).

Cabalistic Hermeticism

Another contribution to the development of Hermeticism came from Giovanni Pico della Mirandola (1463–1494), who added a Christianized cabalistic magic derived from Jewish cabala, which he equated with Lull's Great Art (see Sect. 3.4).[22] Additional interest in cabala was created by the arrival in

[22]Cabala (also cabbala, kabala, kabbalah, etc., from Hebrew *qabālāh*) is a Jewish mystical system, influenced by gnosticism and Neoplatonism, which attempts to find hidden meanings in the Torah through mystical interpretation of the letters of the Hebrew alphabet. (Recall *word magic*, Section 2.1.) It can hardly be coincidental that seven of the nine principles that Lull associates with his figure A (God) are the same as seven of the nine "emanations" (*sephiroth*) from the nameless *ein-soph* (infinity), which together constitute the ten principal names of God (Yates, *OPEA*, p. 12). Also note that the classical text of cabalism (the *Zohar*) was written in Spain in the mid 1270s, which is when Lull was "illuminated" on Mt. Randa. Blau (Cab) gives a brief overview of cabala. Given what we've seen of Leibniz' attempt to construct a calculus of knowledge (Section 4.3), it is hardly

Italy of Jewish scholars after their expulsion from Spain in 1492.

Hermeticism in Copernicus

Finally, we must not ignore more subtle effects of the Hermetic worldview. It is no coincidence that Copernicus' presentation of the heliocentric system in the *Revolutions of the Heavenly Spheres* was written between 1507 and 1530 (published in 1543), in the heyday of Hermeticism, which treated the sun as a "second God." Before his description of the heliocentric theory he cites as authorities Pythagoras and Philolaus, who had defended the heliocentric view in antiquity (1.5, p. 3[b]), and whom we've seen to be successors of Hermes Trismegistus. The Hermetic background is manifest when, after his summary of the theory (1.10), he quotes Hermes Trismegistus:

> In the center of all rests the Sun. For who would place this lamp of a very beautiful temple in another or better place than this wherefrom it can illuminate everything at the same time? As a matter of fact, not unhappily do some call it the lantern; others, the mind and still others, the pilot of the world. Trismegistus calls it a "visible god"; ... And so the Sun, as if resting on a kingly throne, governs the family of stars which wheel around. ... The Earth moreover is fertilized by the Sun and conceives offspring every year. (Copernicus, *De revolutionibus orbium caelestium* 1.10, p. 9[b], tr. Charles Glenn Wallis)

Of course, Copernicus supported his views with mathematical arguments, but that does not decrease the role of Hermeticism in influencing his thought or in the acceptance of his ideas. For example, when Giordano Bruno defended Copernicanism at Oxford, he used citations from Ficino.

5.2.3 Alchemy

> Natural magic is ... nothing but the chief power of all the natural sciences ... — perfection of Natural Philosophy and ... the active part of the same.
>
> — Agrippa (*De occulta philosophia*, Ch. 46)

surprising that he had an active interest in cabala, and in 1687 consulted with Christian Knorr von Rosenroth (1636–1689), a leading cabalist of the time (Scholem, *Kab*, p. 417). Many cabalistic works were placed on the *Index of Prohibited Books* by The Council of Trent (1545–1563); see Yates (*RE*, p. 227).

Natura non nisi parendo vincitur.
(Nature to be commanded must be obeyed.)

— Francis Bacon

Vere scire est per causas scire.
(To know truly is to know through causes.)

— Francis Bacon

That w^{ch} is below is like that w^{ch} is above & that w^{ch} is above is like y^{t} w^{ch} is below to do y^{e} miracles of one only thing

— Newton's translation of the *Emerald Tablet* of Hermes Trismegistus, c. 1690

Dee and Rosicrucian Hermeticism

A third impulse toward the scientific revolution was alchemy, especially combined with Renaissance Hermeticism, as in in the works of Henry Cornelius Agrippa of Nettisheim (1486–1535), Paracelsus (c. 1493–1541) and John Dee (1527–1608).[23] Dee[24] is especially important, since his synthesis of "Magia, Cabala and Alchymia" influenced the Rosicrucian manifestos, which appeared in the very year that the antiquity of the *Hermetica* was disproved, and so inaugurated the "Rosicrucian Enlightenment," the next phase of the Hermetic tradition.[25] The Rosicrucian manifestos helped define

[23]Sources include Clulee (*JDNP*), Yates (*RE*) and Yates (*OPEA*).

[24]Dee was also a devoted Lullian, and had more works of Lull than of any other author in his library, which was one of the largest of the time (about 4000 works). It was assembled as part of his project to prevent the destruction of England's libraries — nearly all of Oxford's was sold — following the dissolution of the monastaries (1536–1539) during the English Reformation. The library occupied many rooms at Mortlake, Dee's estate, which also housed scientific instruments and laboratories, and amounted to a research institute; for two decades it was frequented by many Elizabethan intellectuals. Mortlake was sacked in 1583 by a mob intolerant of Dee's "forbidden knowledge"; approximately 1000 of the works in the library, including many ancient documents, were destroyed. See French (*JD*, pp. 40–61, 204).

[25]The Rosicrucian manifestos (*Fama fraternitatis*, 1614; *Confessio rosae crucis*, 1615), which describe a "Fraternity of the Rosy Cross," may have been written by Johann Valentin Andreae (1586–1654), a Lutheran theologian; he definitely wrote the third Rosi-

the image of modern science by describing a fraternity of learned magicians meeting yearly to share their discoveries, and by announcing the dawn of a new age, in which this knowledge would be used for the benefit of mankind.[26]

Bacon's Rosicrucianism

This view of science is essentially the same as that advanced by Dee's contemporary, Francis Bacon (1561–1626), who is remembered for advocating the experimental method, the formation of learned societies, and the use of science for the good of humanity. It is less well known that he recommended the study of alchemy, natural magic and enchantment ("fascination"), as well as the reform of astrology. However, as a member of the Jacobean court he was obliged to disassociate himself from occult studies, since James I was terrified of magic and witchcraft, and had recommended in his *Demonology* (1597) that witches be executed.[27] Bacon's prudent avoidance of occult subjects may also account for several of his worst failures of foresight, including his rejection of Copernicanism and of William Gilbert's investigations of magnetism ("occult forces"). It also explains his neglect of the role of mathematics in science, since in the seventeenth century "mathematics was still suspected of being one of the black arts… and to ordinary people

crucian document, *The Chemical Wedding of Christian Rosencreutz* (1616). These claim that the Fraternity is several hundred years old, and is — of course — based on ancient wisdom. Kepler is rumored to have been involved in Rosicrucian groups; he certainly knew Andreae — both had been mathematics students of Michael Maestlin (fl. 1590) at the University of Tübingen. There is also a persistent rumor that Leibniz' first job after earning his doctorate in 1666 was as secretary of a Rosicrucian society in Nuremburg; in any case he later used Rosicrucian ideas. See Yates (*RE*, pp. 154–155, 223) and Edwards (*EP*, s.v. 'Leibniz'); translations of the manifestos are in Yates (*RE*, Appendix). The connections between the seventeenth-century Rosicrucian movement and more recent Rosicrucian and Rosicrucian-inspired organizations are tangled and tenuous at best.

[26]Many reformers imagined that this "new age of enlightenment" had begun with the 1613 marriage of Frederick, the Elector Palatine, and the Princess Elizabeth, daughter of James I. They hoped that Frederick would lead the Protestant resistance against the Hapsburg Empire. Andreae was aware of a numerological prediction that 1620 would bring about the downfall of the Pope and Mahomet, and the millennium was expected about 1623. These predictions may have encouraged Frederick in late 1619 to accept the crown of Bohemia when it was offered to him by Bohemian rebels, but he was utterly defeated in 1620, and the Enlightenment was delayed until after of the Thirty Years War (1618–1648). It's apparent that the intellectual movements leading to the scientific revolution depended on the political, religious and cultural context, which is unfortunately beyond the scope of this book; see Yates (*RE*).

[27]Yates (*RE*, p. 123) attributes James's fear to "neuroses about some experiences in his early life."

it was a frightfully dangerous study" (French, *JD*, p. 5).[28] Nevertheless, it's certain that Bacon was influenced by Rosicrucian Hermeticism, for after he died (1626), his manuscript for *The New Atlantis* was found and published; it is a thinly-disguised description of a Rosicrucian utopia.[29] Its description of Salomon's House, the college of the utopia's scientist-priests, was the inspiration for the Royal Society (founded, 1660) and, through it, for all subsequent scientific societies.

Hermetic Alchemy

The purpose of the "new alchemy," which Rosicrucianism added to Hermeticism, was not the physical transformation of base metals into gold, as is commonly supposed, but rather the spiritual transformation of the alchemist (French, *JD*, pp. 76–77). The "prime matter" with which the new alchemists worked was *logos*, the mystic "word" as conceived in the *Hermetica* and the cabala (Yates, *GB&HT*, pp. 150–151; Pagel, *Par*, pp. 82–104, 203–237). Nevertheless, these transformations were treated symbolically, through chemical manipulations, which in addition to laying the foundation of laboratory practice, encouraged the application of mathematics to physical phenomena.

Newton's Alchemical Studies

On the other hand, Isaac Newton (1642–1727) was not especially interested in alchemy as a means of spiritual development, but saw it as a source of insights into science.[30] Although Newton's alchemical studies are often viewed as an embarrassing anomaly in a life otherwise devoted to science, it's more accurate to view the *Principia* and his other scientific work as a digression from a lifetime of alchemical research. For example, over a period of more than twenty years he studied, translated and wrote commentaries on the "Emerald Tablet," the "bible of the alchemists," which was supposed to be an ancient Hermetic document found in the hands of the mummy of Hermes Trismegistus, hidden in the great pyramid of Giza (see quotation, p. 173).

[28] The Greek adjective *mathematikos* originally meant *fond of learning*, then *scientific* and, especially after Aristotle, *mathematical* (*LSJ*). Latin *mathematicus* originally meant mathematician; however, by the first century CE it also meant astrologer (*OLD*). For example, Charlemagne prohibits astrology in these terms: *Ars autem mathematica damnabilis est et interdicta omnino* – The mathematical art, however, is condemnable and entirely forbidden (Lecky, *RIRE*, p. 15, n. 5). "In the thirteenth century Roger Bacon had felt it necessary to distinguish mathematics from divination… In the sixteenth century the word, mathematics, no longer has such a double meaning although it still included astronomy and astrology, so that a royal *mathematicus* might include the drawing up of horoscopes among his functions" (Thorndyke, *HMES*, Vol. 5, p. 14).

[29] The Rosicrucian allusions are made explicit in John Heydon's *Holy Guide* (1662).

[30] Sources include Dobbs (NCET), Dobbs (*FNA*) and Westfall (RANC).

On the basis of Hermetic-Neoplatonic ideas, Newton was convinced that matter is inherently passive, and that the only source of action was the spiritual realm. This active spirit operates on pairs of material opposites (earth/fire, fixed/volatile, etc.) to "procreate a more noble offspring" (Newton, *Commentary*; Dobbs, NCET, p. 184). Thus, although "Newtonian philosophy" is often equated with mechanistic explanations, including the reduction of biology to physics, he himself thought that physical phenomena derive from biological phenomena, specifically, from a "vegetable spirit" that is the source of all activity: "all matter duly formed is attended with signes of life" (Newton, quoted in Dobbs, NCET, p. 185). "Vulgar chemistry," which — in contrast to alchemy — treats matter as though it's inanimate, applies only when the bulk of matter is sufficient to hide the operation of the "vegetable spirit." He thought this spirit was "signified" by Hermes.

It is hardly surprising that Newton's theory of gravity was criticized for introducing "occult forces" into physics, and that he felt obliged to withhold any explanation of the source of the gravitational force with his famous *hypotheses non fingo* (I do not invent hypotheses) — although in the last paragraph of the *General Scholium* to the *Principia* he does attribute activity in the universe to "a most subtle spirit which pervades and lies hid in all gross bodies." Newton's greatest scientific accomplishment was to show — against Aristotle — that the same laws apply in the celestial realm as in the mundane realm, or, in the words of the Emerald Tablet, "that which is below is like that which is above, and that which is above is like that which is below."

5.2.4 Renaissance Magic and Science

> *Magus significat hominem sapientem cum virtute agendi.*
> 'Magician' signifies a wise person with the power to act.
>
> — Giordano Bruno

> It is one of the more pleasant ironies of history that the driving force behind the development of the new physics and astronomy was a semimystical religion.
>
> — W. T. Jones (*HWP*, Vol. 3, p. 93)

tive
ervention in
ture

One might well wonder what this odd mix of mystical doctrines and magical practices could have to do with the scientific revolution. First, it led to the "ennoblement of magic": what had been seen as the invocation of evil spirits, was now seen as the manipulation of natural forces; "the magician was now seen as a pious religious philosopher with insights into the secrets of the divine and natural orders" (Clulee, *JDNP*, p. 5). This encouraged a shift from medieval science's passive contemplation of "God's handiwork" to the active intervention in phenomena that characterizes modern *experimental* science. This was reinforced by the magical tradition, in which trials and tests of ideas were taken for granted. (We must not forget that science in the Greek tradition was almost entirely speculative and nonexperimental.)

Role of Mathematics

Needless to say, the development of Renaissance Hermeticism took place on the background of pervasive Pythagorean-Neoplatonism that we've already discussed. Therefore, the Hermetic magicians' "will to operate" — to work with divine and natural forces — was combined with the belief that these forces obey mathematical laws. This was reinforced by the use of mathematics in alchemy and astrology.

Importance of Practical Arts

The Renaissance magician was also expected to be skilled in many practical arts, including mathematics and mechanics. Dee, for example, was an expert in astronomy, optics, navigation and mechanics. E. G. R. Taylor has called him "a very Atlas bearing upon his shoulder the sole weight of the revival in England of the mathematical arts" (Dee, *MP*, p. 1). In his famous preface to the first English translation of Euclid (1570) Dee argues that mathematics is a prerequisite to all the arts and sciences (magical and otherwise), since "*All thinges ... do appear to be Formed by the reason of Numbers. For this was the principall example or patterne in the minde of the Creator*" (Dee, *MP*, sig. *i^{r}).

Summary

In summary, we find in Renaissance Hermetic-Neoplatonism the source of the two essentials of modern science: the mathematical description of nature, and the intervention in phenomena for empirical tests as well as practical ends. When in 1614 Isaac Casaubon (1559–1614) showed that the *Hermetica* couldn't be any older than the second or third century CE, they lost much of their fascination for the Renaissance mind, with its excessive reverence for antiquity, but their role was quickly filled by the Rosicrucian manifestos. In any case, the mathematico-experimental method was so well established by then that the scientific revolution could roll on, unencombered by its magical past.

This rough magic
I here abjure. . .
I'll break my staff,
Bury it certain fathoms in the earth,
And, deeper than did ever plummet sound,
I'll drown my book.

— Prospero in Shakespeare's *Tempest*, 5.1.50

5.2.5 The Witches' Holocaust

Bodin and the Witch-hunts

The renunciation of magic was just as well for science, since the shadow of the witch-hunts was already over Europe. One cause was the 1580 publication of *De la démonomanie des sorciers* by Jean Bodin (c. 1530–1596), a book — scholarly in tone — that did much to fan the fires of the witch scare in the late sixteenth and early seventeenth centuries.[31] Bodin was a magistrate and his book included detailed prescriptions for the trial and punishment of witches; on the latter question he cited the Bible: witches must die.[32]

Bodin was especially vehement in his condemnation of the Hermeticists Pico and Agrippa, whom he accused of the blackest of black magic. The basis for his accusation was their cabalistic magic; they had claimed that by using the Christian cabala they had ensured that their magic used only "angelic powers," but Bodin argued that all magical (non-contemplative) use of cabala

[31] The premier "witch-hunter's manual," *Malleus maleficarum* (*Witches' Hammer*), had been first published a hundred years before, c. 1487, by two Dominican monks, Sprenger and Krämer, under the direction of Pope Innocent VIII. Sprenger suggested torturing a witch for two or three days, and recommended that the tortures be more painful than usual, since witches were supposed to be able to deaden pain magically (Lecky, *RIRE*, Vol. 1, p. 23b, n. 3).

[32] Exodus 22: 18: *Maleficos non patieris vivere*. This is usually translated, "You shall not allow witches to live," but the word *maleficus* originally meant evil-doer or criminal, not witch; the usual Latin words for witch are *maga*, *venefica* and *saga* — which also means wise-woman (*OLD* s.vv.). However, the Hebrew word that appears in this passage is feminine, and usually refers to a witch or sorceress; it probably derives from an Akkadian root meaning to cut or mix, and alludes to the use of herbs for healing, poisoning and magic (Cassuto, *CE*, p. 95; Hastings, *DB*, p. 608; Douglas, *NBD*, p. 766). Childs (*Ex*, pp. 477–478) notes that the phrase used was not the usual one for the death penalty, but probably meant banishment (i.e., you will not allow witches to live among you). It is worthwhile to consider how many women's lives were affected by the translation of a word or two. Cf. also Lev. 20: 27, Deut. 18: 10f.

was demonic. Further, he claimed that the supposed spread of witchcraft was a result of possession by demons released through the Renaissance magicians' use of cabala.

Execution of Giordano Bruno

One victim of Bodin's campaign was Giordano Bruno, who along with others hoped that Hermeticism would be a "religion of natural contemplation" that might heal the strife in Europe resulting from the Reformation (1517) and Counter-Reformation (1560). Accused of sorcery, he was held in the Inquisition's dungeons for eight years, during which time he was frequently interrogated, and was finally convicted in 1600. They stripped him naked, staked his tongue so that he couldn't "speak blasphemies," and burned him alive at the stake (Durant, *Civ 7*, pp. 621–624).[33]

Magi and Witches Treated Differently

Most of the victims of the witch-hunts were elderly women, "village witches," who probably had never heard of Pythagoras, Hermes Trismegistus or cabala.[34] Yet during the peak of the witch craze as many as 100 supposed witches were burned daily in the carnival-like *auto-da-fé*, and before it ended many thousands, mostly women, had perished; perhaps 100 000 in Germany alone — the Witches' Holocaust (Bynum & al., *DHS*, p. 447; Durant, *Civ 7*, p. 578). Some people, including Agrippa and his student Johann Weyer (fl. 1563), courageously defended the women, but these so-called "hag-advocates" came under nearly as vehement attack as the witches themselves, because they undermined belief in witches; eventually it was considered heretical to *disbelieve* in witchcraft![35]

The Death of the Renaissance

Although relatively few intellectuals were tortured or executed (though even Henry III was accused), the message was clear and Renaissance magic was effectively suppressed. Marlowe's Faust replaced Shakespeare's Prospero as the image of the magus. Throughout Europe, in the words of Yates

[33]The association between heliocentricity and Hermeticism, exhibited by Bruno and Copernicus, may have contributed to the Inquisition's 1616 declaration that heliocentricity is heretical and to its 1633 condemnation of Galileo. In fact, Cardinal Bellarmine, the Church's chief theologian, who had framed the decision against Bruno, had Galileo under surveillance, for he suspected a new outbreak of Pythagorean "novelties." Also, Galileo's well-known statement (1601), "The book [of the universe] is written in the mathematical language," no doubt sounded suspiciously Pythagorean to the watchful Cardinal. See De Santillana (*CoG*, pp. 28–29) and Yates (*GB&HT*, pp. 358–359). In 1922 the Church finally abandoned its official rejection of heliocentricity.

[34]For example, "while Johannes Kepler's mother was pursued by the authorities as a witch, Kepler himself, despite the many mystical, spiritual, magical, and even heretical parts of his scientific work, was never seriously threatened (Estes, GW, p. 163).

[35]See Wiener (*DHI*, Vol. 4, pp. 521a–523a) and Bynum & al. (*DHS*, p. 447).

(*OPEA*, p. 92), "the Renaissance went down in the darkness of the witch-hunts."[36]

> It is a heretic that makes the fire,
> Not she which burns in 't
>
> If this be magic, let it be an art
> Lawful as eating.
>
> — Shakespeare, *Winter's Tale*, 2.3.114, 5.3.77

5.2.6 Belief and the Practice of Science

> . . . when they come to model Heaven
> And calculate the stars, how they will wield
> The mighty frame; and build, unbuild, contrive
> To save appearances; how gird the sphere
> with centric and eccentric scribbled o'er,
> Cycle and epicycle, orb in orb:
>
> — Milton, *Paradise Lost*, 8

Condemnation of Aristotelianism

You may have been shocked to discover that the heroes of the scientific revolution were firmly committed to mystical and magical beliefs discredited by contemporary science. I've discussed some of the reasons — the use of mathematics and the will to operate — but there is another that it will be valuable to consider briefly: the role of belief in the practice of science. To understand this, we must look again at the late twelfth century, which gave birth to the first universities, at Paris, Bologna and Oxford.[37] Aristotle's works in natural philosophy had just become available in Latin translation, and Aristotelian logic and science formed the core of the curriculum for all students, whether of law, medicine or theology. The trouble was that Aristotle's philosophy differed from medieval theology on a number of points; for example, Aristotle argued that nature follows regular and immutable laws, which implied the impossibility of miracles, upon the existence of which

[36]She also notes, "witch-hunts can always be used against personalities or politico-religious movements which the hunters desire to eliminate, concealing them — even from future historians — within folds of diabolic propaganda" (Yates, *OPEA*, p. 71).

[37]Sources include Grant (*PSMA*, Chs. 3, 4, 6).

Christianity depended. The exposure of these differences led to a series of condemnations of Aristotelianism: in 1210 reading of Aristotle — in public or in private — was forbidden in Paris, a prohibition that was repeated, extended and strengthened in 1215, 1231 and 1245. In 1270 the Bishop of Paris threatened excommunication against anyone accepting certain Aristotelian propositions. The climax came in 1277, when Pope John XXI ordered an investigation of the University of Paris, which resulted in the condemnation of 219 Aristotelian propositions, including some advocated by Aquinas. The penalty for defending any of them was automatic excommunication.

Liberating Effect of Condemnation

Despite its repression of free inquiry, the condemnation of 1277 had a liberating effect on medieval science, since by undermining the authority of Aristotle it opened the way for new methods. And indeed in the early fourteenth century we find Ockham espousing a radical empiricism destined to lead to modern science. Nevertheless, the scientific revolution began in the sixteenth century, not in the fourteenth, so we must ask what other other factors were involved.

Double Truth

Singled out for condemnation in 1277 was the "doctrine of double truth," the view, apparently held by a number of the Paris faculty, that there were separate truths in science and religion. Thus certain propositions, such as the eternity of the universe, could be accepted as irrefutably demonstrated in Aristotelian science, while exactly contradictory propositions, such as that the universe was created, could be accepted as revealed truth. With this means of reconciling the contradictions of reason and faith closed to them, medieval scholars needed a more subtle way of studying science while preserving orthodoxy, which they found in a kind of "Christian positivism."

Positivism and Saving the Phenomena

Positivism, a term popularized by Auguste Comte (1798–1857), refers to the view that one should not make metaphysical claims beyond the existence of what is given: sense data or empirical observations. Since at least the Hellenistic period, it was known that there are many ways to "save the phenomena," that is, that many different hypotheses can explain the same empirical data.[38] For example, the Ptolemaic system of epicycles can "save the phenomena" — explain the observed motion of the planets — as well as the heliocentric system of Copernicus. Indeed, defenders of Copernicanism

[38]The verb conventionally translated "to save" is σῴζειν, and in this context means to preserve, maintain or observe (*LSJ*). The recognition that there are many ways to save the phenomena (i.e., appearances) may go back to the second century BCE astronomer Hipparchus (Bochner, *RMRS*, p. 123).

(though not Copernicus himself) claimed that his theory was not heresy, since it did not require one to believe the earth *really* moves around the sun, but that the fiction that it does makes calculation easier. In a similar way, scientists in the fourteenth century were able to continue their investigations and avoid excommunication by claiming their theories were merely convenient ways of summarizing observations, and by avoiding any claim that they were really *true* (only the Church could pronounce on metaphysical truth).

Secundum Imaginationem

A valuable weapon in this defense was the phrase *secundum imaginationem*, "according to the imagination." Thus, continuing the heliocentric example, a scholar might say, "Imagine that the earth moves around the sun — of course I don't believe such heresies — but if we imagine it does, then so and so and such and such follows." This approach had an enormously liberating effect on scientific theorizing, since it allowed scientists to entertain hypotheses "for the sake of argument," and to follow out their logical consequences, regardless of whether they were heretical or even contrary to common sense. For example, medieval physicists considered the consequences implied by infinite velocities and perfect vacuums.

Disadvantages of Positivism

Since the condemnation of 1277 weakened Aristotle's authority, and the positivist approach allowed new and potentially heretical hypotheses to be entertained, we must ask again why the scientific revolution began in the sixteenth century rather than the fourteenth. We have seen part of the answer: mathematical and experimental methods were encouraged by the rebirth of Hermeticism, which followed the return of the *Hermetica* to the West in 1460. The rest of the answer is that positivism subtly discouraged the new science.

The fourteenth-century scientists held that only the revealed truths of religion could provide any certainty about the universe; science could, at best, tell plausible stories. Therefore, the theories they proposed were not taken very seriously, by the scientists themselves or by others. Even if they believed that they were true, the threat of excommunication forced them to express the theories in hypothetical language. Although they anticipated many important mathematical and scientific developments, the general pattern is that these good starts were not pursued. Such can be the stultifying effect of positivism.

Copernicus Believed his Theory

Copernicus was able to go beyond his contemporaries because he wasn't a positivist; as we've seen, he was influenced by Hermetic Neopythagorean ideas, and so he believed that his mathematics truly described the structure of the universe. Copernicus may have been saved from Church persecution by Andreas Osiander (1498–1552), a Lutheran theologian, who — without Copernicus' permission — inserted a forward into the first edition of *De revolutionibus* stating that astronomers "cannot by any line of reasoning reach the true causes of these movements . . . , but it is enough if they provide a calculus which fits the observations" (p. i^b^). Other scientists, however, especially Kepler, who found the idea of occult forces radiating from the sun quite compatible with their Hermeticism, interpreted Copernicus' theory to be an account of the actual structure of the universe.[39] This was important for the development of science, since if the planets actually moved as Copernicus said, then the laws of motion would have to be different from what Aristotle had proposed. Thus Copernicus set in motion the new investigation of the laws of motion that would lead to the work of Galileo and Newton.

Importance of Emotional Commitment

It is a commonplace that scientists should be objective about their investigations, and it is often recommended that they maintain multiple hypotheses so that they will balance their emotional biases. Time and again, however, the greatest discoveries have been made by scientists who were fervent believers in their worldviews — regardless of whether those worldviews were correct (in the current view). It seems that only belief can provide the emotional motivation to sustain long-term research against prevailing opinion. Nobody is going to stake their career — or their life — on a theory that they think is merely plausible, or a handy rule for calculation.

The obvious danger is that such emotional commitment can blind the researcher to contrary evidence. The successes are remembered as the heroes of science, who had the courage of their convictions; but the failures are remembered as quacks (J.B. Rhine, I. Velikovsky), or as great scientists whose conviction turned to obsession, who couldn't give up a pet idea (Einstein: general field theory), who couldn't see the handwriting on the wall (Einstein: quantum mechanics), or who perhaps became senile or even crazy in their old age (Newton: alchemy). The practical problem for all scholars is to know when to persist in spite of obstacles, and when to abandon a lost cause. The factors affecting the rise and decline of scientific theories — the dynamics of knowledge — will be a recurring topic in the rest of this book.

---

[39] At first Kepler — wisely — hid his Copernicanism: in his student days (1590) he defended it as a "devil's advocate"; in 1596 he almost let the cat out of the bag in his *Cosmic Mystery*, but the Rector of Tübingen warned him to present his ideas "only as a mathematician" and not to claim they described the actual nature of things. His advocacy of heliocentrism became open in *The New Causal Astronomy* (1609) and *The Musical Theory of the World* (1619). See Wightman (*SiRS*, Chs. 9, 11).

5.3 Reduction of Continuous to Discrete

5.3.1 The Problem of Motion

The Irrationality of Change

In Chapter 2 we saw that the Pythagoreans associated *change* with the indefinite and undetermined, the irrational and evil. Some of the Greeks held change to be illusory — Parmenides said the universe is "changeless within the limits of great bonds, without beginning or ceasing" and that "Fate fettered it to be whole and changeless"; Zeno's paradoxes were probably intended to support this doctrine. Plato held that true knowledge was possible only of the changeless world of forms, and that the natural world, where change is pervasive, is ultimately unintelligible. Aristotle rejected this and devoted much of his work to change, but his theory of change was almost entirely qualitative. He had reduced change to *logoi* in one sense — words — but not in the mathematical sense: ratios. The goal of the Pythagorean program — the reduction of nature to *calculation*, to *figures* — had not been achieved.[40]

This is still true today: many of the phenomena with which Aristotle dealt, such as biological development, meteorology and the dynamics of thought, have not been described adequately by mathematical theories; a fully *rational* account eludes us. Nevertheless, one of the great accomplishments of the scientific revolution was the rationalization of *motion*, literally, its reduction to ratios. This is usually credited to Galileo, but we will see that his work was the culmination of a gradual application of mathematics to change.

Aristotle on Velocity

The idea of *instantaneous velocity* is fundamental to modern mechanics, yet Zeno had already shown deep problems in the concept (Section 2.3), and Aristotle had rejected it outright: "nothing can move in *the now* (τὸ νῦν) . . . , but truly nothing can be at rest" in it either (*Phys.* 234a24, 33); for him there were only *average* velocities.[41] The mathematical analysis of motion was further complicated by the medieval view that velocity is not a quantity. We have already seen part of the reason for this view: ratios of like magnitudes

[40]The quotes from Parmenides are DK28b8 or *KRS* 298, 299. We'll see in Ch. 10 that another Greek philosopher, Heraclitus, faced up to the pervasiveness of change, though he couldn't reduce it to mathematical ratios.

[41]Sources for this section include Clagett (*SMMA*, pp. 199–219), Clagett (RS), Maier (*OTES*, Chs. 1, 7), Grant (*PSMA*, Chs. 4,6) and Dijksterhuis (*MWP*, II.v.C). Clagett (*SMMA*) contains much source material in translation.

were not considered numbers, and ratios of unlike magnitudes (such as a ratio of a distance and a time) were not recognized at all. Understanding the full reason, however, requires some discussion of the Aristotelian distinction between *quantity* and *quality*.[42]

Quantity

Quantity, as one would expect, refers to the amount of something, either a *discrete multitude* or a *continuous magnitude*. In modern terms, they are the sorts of things that are obviously measurable by numbers. Discrete multitudes are measured by counting; continuous magnitudes are measured with respect to a unit of the same kind (an area by an area, a time by a time, etc.). In the middle ages the terms *augmentation* and *diminution* were used to refer to the increase and decrease of a quantity, the significant point being that a quantity of some kind is increased through augmentation by adding more of the same kind.

Quality

A *quality* is a property or determination of something, which is often variable in *degree* (*gradus*) or *intensity*, but not in amount. Consider color; we know it can vary continuously from red to purple. But what is there a greater (or lesser) amount of in red than in purple? The modern answer is wavelength (or frequency), but that was not known in the middle ages. Other qualities are hardness, roughness, hotness, consonance, health, courage, wisdom. These are harder to think of as quantities, since we have difficulty answering, "Quantity of what?"

Intension and Remission of Qualities

One of the distinctions between quality and quantity is that although both can vary, they do not do so in the same way. We cannot "increase" the degree of a quality by adding more of the same: We cannot get a red object by combining enough blue objects, or get a hot thing by combining many cool things, or get a smooth surface from many rough surfaces, etc. — the cooperation of many fools yields not wisdom. For this reason, we do not speak of the augmentation and diminution of qualities, but of their *intension* (*intensio*) and *remission* (*remissio*), collectively referred to as *alteration*.[43] Science must study change, and since many of the changes in nature are

[42] These are two of Aristotle's famous *categories*; in a typical list (*Cat.* 4.1b25–2a4), the others are substance, relation, action, passion, place, time, position and state. These correspond to the kinds of questions in Ancient Greek. Medieval use of quality and quantity is summarized in McKeon (*SMP*, Vol. 2, Gloss., s.vv. 'alteratio', 'augmentio', 'motus', 'qualitas', 'quantitas'). See also Bochner (*RMRS*, pp. 226–229).

[43] Note that this use of *intension*, and the later use of *intensive* and *extensive* is unrelated to the *intension* and *extension* discussed in Section 4.3.2 (p. 119). It is of course also distinct from *intention* (with a 't', p. 71).

alterations of degree or intensity, a major goal of thirteenth-century science was to explain the intension and remission of qualities.

Location as a Quality

Now we return to the problem of motion, for in the middle ages location was considered a *quality*. This may seem peculiar, but it is quite reasonable if we recall: (1) like color — clearly a quality — location can vary in several different dimensions; (2) location cannot be a quantity, for what is it a quantity of? (Recall that representation of location by coordinates comes much later, Section 5.1.2.) In this view, motion then is a kind of *alteration* of location.

Velocity as a Quality

A more sophisticated medieval view — one which dealt better with Zeno's paradoxes and helped the advance of physics — was to view *velocity* as a quality.[44] It certainly seems that at each instant of time the arrow at rest is qualitatively different from the arrow in motion, and by extension there is a difference of intensity of this quality in the swiftly and slowly moving arrows. What we call the instantaneous velocity of the arrow was for them a quality, *velocitas*, which the arrow has to a varying degree at different times. The variation of the arrow's speed in time is analogous to the variation of a slide-whistle's pitch over time, or the variation over distance of the temperature of a metal bar heated at one end. By this view, intension and remission of velocity correspond to acceleration and deceleration.

Intensive vs. Extensive Quantities

Towards the end of the thirteenth century many explanations were proposed for the intension and remission of qualities, but the one relevant here viewed the intensity of a quality as a quantity. Already by Aquinas' time (mid-thirteenth century) a distinction was being made between the intensive (or "virtual") quantity of a quality and the usual extensive (also "dimensive," "corporeal") quantities. For example, the schoolmen distinguished the intensity of heat (modern temperature) — an intensive quantity — from the total heat in an object. Similarly, they distinguished intensive weight (modern specific weight) from extensive weight (modern total weight), a distinction analogous to that between density, an intensive quantity, and mass, an extensive.

Quantification of Degree

Richard Middleton may have been the first (c. 1281) to state that intensive quantities, like extensive ones, can be increased through addition, a view which probably was supported by Duns Scotus and certainly was by Scotus' student Joannes de Bassolis. During the fourteenth century, there was a rush to quantify various qualities, including many which still resist quantification,

[44] Of course, the notion of a velocity *vector* would not be invented for many centuries.

such as charity, sin and grace. The impetus for this quantifying frenzy may have been the first translations into Latin of some important ancient texts. Aquinas, who was dissatisfied with existing Aristotelian texts translated from Arabic, had encouraged his friend William of Moerbeke (c. 1215–c. 1286) to translate them from the Greek. The resulting translations include Aristotle's *Categories*, which discusses quantity and quality, and the *Commentary on the Categories* by the sixth-century Neoplatonist Simplicius, which forms a foundation for much of the scholastic discussion of intension and remission. In 1269 Moerbeke also translated almost all of the works of Archimedes, whose practical application of mathematics to physical problems may have also encouraged attempts to quantify qualities. Also we should not ignore the Neoplatonic background of all contemporary philosophy, which would have been strengthened by the 1277 condemnation of Aristotelianism.

Natural Philosophy at Oxford and Paris

The break from Aristotelianism and founding of the new science was largely the work of Thomas Bradwardine (c. 1290–1349) and his colleagues at Oxford University, and of Jean Buridan (c. 1295–c. 1358) and his colleagues at the University of Paris.[45] I'll begin with Bradwardine's group at Merton College, Oxford, since they seem to predate slightly the Paris group (Thorndyke, *HMES*, Vol. 3, p. 374).

Criticism of Aristotle

In his *Physics* and *On the Heavens* Aristotle had given quantitative laws of motion, which related the magnitudes time, distance, "force," and "resistance"; naturally, these laws did not deal directly with velocity, which was a ratio, not a magnitude. These laws were criticized by the fourteenth century schoolmen, largely because they misunderstood what Aristotle had said.[46] Nevertheless, this criticism further weakened the Aristotelian hegemony and encouraged exploration of alternative theories of motion — an example of how even misunderstandings may serve to advance knowledge.

The Calculators of Merton College

Although the thirteenth-century geometer Gerard of Brussels may have been the first to treat velocity as a magnitude, the major impetus in this direction came from the fourteenth-century proclivity to treat intensities as quantities. This was especially true of the Mertonians, who were collectively called "The Calculators," and who developed the method of Calculations. This involved the writing of expressions "in terms," that is, using letters to stand for physical magnitudes. At first, this was simply a convenient short-

[45] Recall Buridan's investigation of logical paradoxes, pp. 75, 76.

[46] Specifically, they thought his laws implied that a force, however small, could move a body, however large, but Aristotle clearly said that these laws no longer apply when the force is insufficient to set the body in motion.

hand, but later it developed into a precursor of algebra, a kind of calculus that encouraged *calculating* with physical magnitudes.

Investigations at Merton College

Much of the Calculators' effort was devoted to understanding the continuum, and especially to the problem of describing continuous change in terms of ratios and proportions. For example, Bradwardine wrote the *Treatise on the Proportions of Velocities in Movements* (1328) and *On the Continuum*. The foremost Mertonian, Richard Swineshead, known simply as The Calculator, wrote *On Motion* and the *Book of Calculations* (c. 1346), which dealt perceptively with infinities and infinitesmals. Other natural philosophers working at Merton college from 1328 to about 1350 were John Dumbleton, and William Heytesbury, who proved the Uniform Acceleration Theorem (experimentally confirmed by Galileo) and attempted to define instantaneous velocity. Their arguments foreshadowed many of those used by nineteenth-century mathematicians to justify the infinitesmal calculus.

Configuration of Forms

Nicole Oresme (c. 1323–1382) was one of the most important figures in Buridan's group at Paris, who were called the *Terminists*; I have already mentioned the role his diagrams had in the development of analytic geometry (p. 156), but they also were important in the mathematization of motion. Oresme's diagrams show the relation between intensive and extensive quantities in an object (Fig. 5.4). The vertical axis, called *longitude*, represents the intensity of a quality, and the horizontal axis, called *latitude*, represents some dimension over which the intensity varies; it could be space or time. Thus Fig. 5.4B could represent the temperature of a metal bar heated at its left end, or the descending pitch of a slide-whistle. The *configuration* (or shape) of the curve of intension was supposed to determine the faculties of animals and plants, the "occult properties" of herbs and stones, and the properties of chemical compounds.

Of course Oresme's diagrams stimulated the graphing of physical quantities, but their significance for the development of mechanics lay in the treatment of velocity as a quality. Since velocity was the *intensity* of a motion, it was represented by the longitude or height of the intension curve. For example, Fig. 5.4B could represent a uniformly decelerating motion. The diagrams encouraged viewing velocity as a magnitude, and gave reality to the idea of instantaneous velocity, since it was represented by longitudes in the diagram (e.g., XY in Fig. 5.4B).

Uniform Acceleration Theorem

Oresme's diagrams simplified conceptualizing and proving theorems about motion. For example, the Uniform Acceleration theorem, "probably the most outstanding single medieval contribution to the history of physics" (Grant,

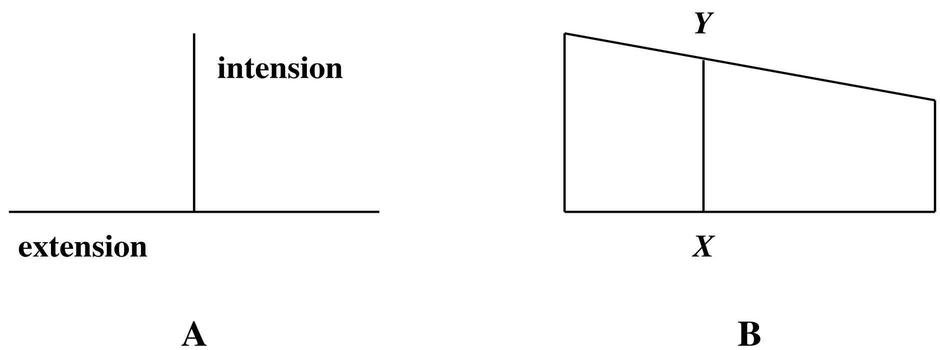


Figure 5.4: Oresme's diagrams of the quantity of a quality. **(A)** The vertical axis (longitude) represents the intensity or degree of a quality; the horizontal axis (latitude) represents a dimension (e.g. space or time) over which the intensity varies. **(B)** This diagram represents a uniform decrease, from left to right, of the intensity of a quality. It could represent the temperature of a metal bar heated on the left end, the decreasing pitch of a slide-whistle, or the velocity of a uniformly decelerating body. The longitude XY represents an intensity at a particular point in the latitude, for example, the temperature at a point, or an instantaneous pitch or velocity. The shape of the "intension curve" was called the *configuration* of the quantity, which was thought to determine the properties of the quality.

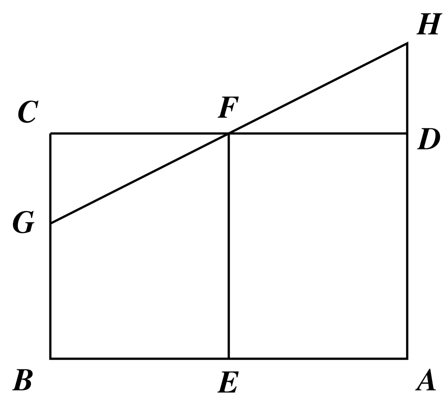


Figure 5.5: Demonstration of Uniform Acceleration Theorem. GH represents uniformly accelerating motion starting with velocity BG and ending with velocity AH. E bisects AB and so CD represents constant motion with velocity EF equal to the average of initial and final velocities BG and AH. Triangles CFG and DFH have the same area, so the areas under CD and GH are equal, and so also are the distances covered by the two motions.

PSMA, p. 56), which was proved by the Mertonians, is easily demonstrated by Fig. 5.5. Since latitude represents time and longitude represents velocity, the area under an intension line represents the distance traversed by the motion. The line GH represents a uniformly accelerating motion starting with velocity BG and ending with velocity AH. The line CD represents a constant motion with velocity EF equal to the average of the initial and final velocities BG and AH (since E bisects AB). Since triangles CFG and DFH obviously have the same area, the areas under CD and GH are the same, and so also are the distances covered by the two motions. Galileo uses essentially this diagram in his proof of the theorem in the *Discourse on Two New Sciences*.

5.3.2 Berkeley: Critique of Infinitesmals

> [Calculus:] the art of numbering and measuring exactly a Thing whose Existence cannot be conceived.
>
> — Voltaire

> And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?
>
> — Berkeley, *The Analyst*

Problems Motivating the Calculus

According to Kline there were four major problems motivating the development of the infinitesmal calculus in the seventeenth century: instantaneous velocity, tangents to curves, maxima and minima of curves, and lengths, areas and volumes determined by curves.[47] Although mathematicians of the time saw them as distinct problems, they all involve the "elementary constituents" of curves, conceived either as points or infinitesmal line segments (Fig. 5.6). Instantaneous velocity is apparently the division of an infinitesmal distance by an infinitesmal time, but not a division of 0 by 0. The tangent is determined by two points on a curve that are infinitesmally close together. At a maximum or minimum of a curve infinitesmally close points have the same height. The length of a curve is obtained by adding the lengths of an infinity of infinitesmal straight line segments. And so forth. These were critical problems for mathematical physics, and during the seventeenth century dozens of mathematicians worked on them.

Two Basic Methods

Of course all the mathematicians working on the calculus were well aware of the Greek method of exhaustion (Section 2.6.3); the problem was to find a general method for dealing with infinitesmals. The solutions proposed can be divided into two broad classes, that I will call the methods of *indivisibles* and *evanescents*.

Method of Indivisibles

The idea behind the *method of indivisibles* comes straight out from Pythagoras: a line is composed of an infinite number of indivisible units, as a plane is of lines, and a volume of planes. For example, if the area under a curve is given approximately by the areas of a finite number of rectangles of nonzero width, then its area will be given exactly by an infinite number of zero-width rectangles (Fig. 5.7). Of course a zero-width rectangle looks a lot like a line, and so they were treated. Zeno had already pointed out the problem with this: it is hardly obvious that you can get a "something" by adding together an infinity of nothings (Section 2.3.2). Nevertheless, the method of indivisibles could be used, with care, to solve many problems. Kepler used

[47]Sources for this section include Kline (*MT*, Chs. 17, 19, 40). For selections from Berkeley's *Analyst* see Newman (*WM*, pp. 288–293) and Struik (*SBM*, pp. 333–338).

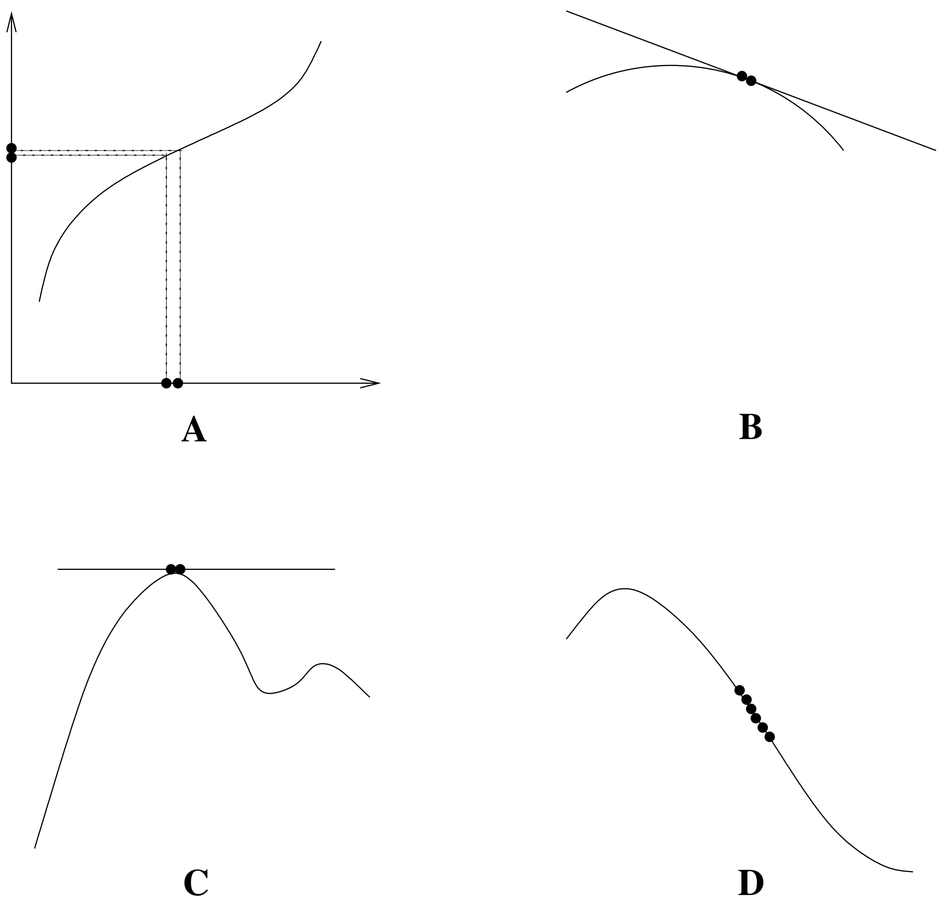


Figure 5.6: Four Problems Motivating Infinitesmal Calculus. The four problems motivating the calculus in the seventeenth century all involved the infinitesmal components of curves. In each case the two dots portray two infinitely close points bounding an infinitesmal line segment. See text for explanation.

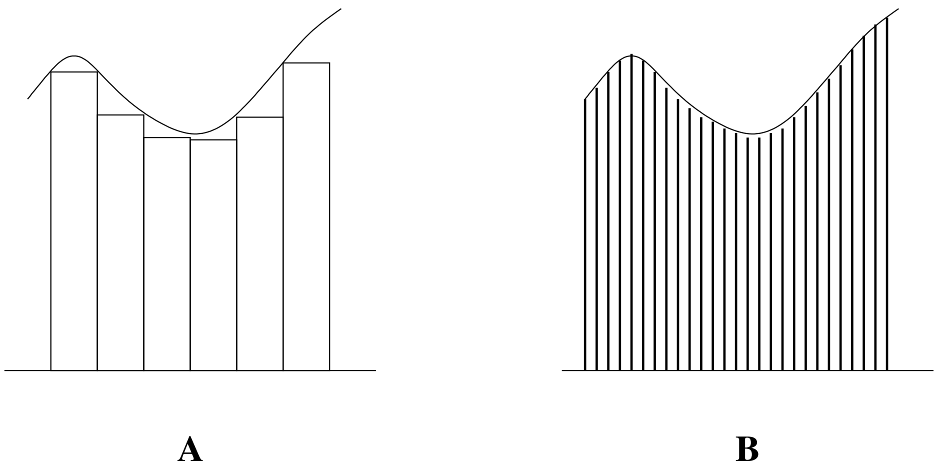


Figure 5.7: Method of Indivisibles. If the area under a curve is approximated by a finite number of rectangles of nonzero width (A), then it will be given exactly by an infinite number of zero-width rectangles (B).

it informally to find the areas of ellipses and volumes of wine casks, as did Galileo to prove the Uniform Acceleration Theorem and other results. It was Galileo's pupil, Bonaventura Cavalieri (1598–1647), who in 1635 first presented the method systematically, but it lacked rigor and didn't convince critics.[48]

A number of mathematicians, including Fermat and Pascal, recognized that an infinitesmal rectangle is not the same thing as a line segment, and that this could lead to errors in the method of indivisibles. This is illustrated by the tangent to a curve, which can be approximated by the hypotenuse of a triangle at that point (Fig. 5.8A), and so is given exactly by the infinitesmal *characteristic triangle* (Fig. 5.8B). The method if far from clear, however, since it's difficult to say how characteristic triangles differ from one another, since they all seem to be identical to points. Fermat and Pascal tried to evade the difficulties by claiming that the calculus was no more than a shorthand for the method of exhaustion, but they were never specific about the connection.

Leibniz tried to avoid the problem by calling the characteristic triangle a "form without a magnitude." He said that infinitesmal forms such as this were smaller than any nonzero magnitude but not zero — certainly

[48]Cavalieri wrote on mathematics, including logarithms, conics and trigonometry, and on its applications, including optics, astronomy and astrology (Smith, *HM*, Vol. 1, p. 362).

A **B**

Figure 5.8: Characteristic Triangle. **(A)** The slope of the tangent at a point is approximated by the hypotenuse of a triangle at that point. **(B)** The exact slope is given by the infinitesmal characteristic triangle, which has shape but no size.

a questionable notion. At other times he claimed that infinitesmals, like imaginary roots, are just formal tools that can in principle be eliminated, but in truth he wasn't very worried about them since they fit into his philosophy of ultimate particles, the *monads*.

Method of Evanescents

Newton and other mathematicians rejected the notion of indivisibles, and agreed with Aristotle in taking infinite divisibility to be the essence of a continuum (see quote, p. 204). Yet if we imagine a triangle such as Fig. 5.8A shrinking to zero, it seems that the hypotenuse must coincide with the tangent just when the triangle shrinks to nothing. Therefore, the slope would be given by the ratio of its just-vanishing (*evanescent*) legs — hence the *method of evanescents*. It was imagined that at just the moment when the triangle vanished, it would retain its shape but have no size. Alternately, infinitesmals were explained as *incipient* or *nascent* (just coming to be) quantities. In discussing the definition of a derivative, which he called a *fluxion*, Newton said:

> by the ultimate ratio of evanescent quantities is to be understood the ratio of quantities, not before they vanish, nor afterwards, but with which they vanish. In like manner the first ratio of nascent quantities is that with which they begin to be. (*Principia*, 1.2, Scholium to Lemma 11)

Although the concept of a limit is apparent here, the flurry of fancy terminology could not disguise the absence of a clear definition.

Ghosts of Departed Quantities

Most mathematicians were not especially worried about the lack of rigorous foundations for the infinitesmal calculus, and were finally embarrassed into putting their house in order by a 1734 essay by Bishop George Berkeley (1685–1753), called *The analyst, or a discourse addressed to an infidel mathematician*.[49] Berkeley was distressed by the scientific revolution and criticized mathematicians and philosophers who claimed that the new materialistic theories of the universe were clearer and more certain than religious doctrine:

> That men who have been conversant only about clear points should with difficulty admit obscure ones might not seem altogether unaccountable. But he who can digest a second or third fluxion, a second or third difference, need not, methinks, be squeamish about any points in divinity

He correctly pointed out that verbiage like "ultimate ratios of evanescent quantities" explains nothing, since the quantities are either nonzero or zero, and when nonzero their ratio doesn't give the exact derivative, but when zero the ratio isn't even defined. It was the middle of the nineteenth century before mathematics and physics were exorcised of the ghosts of departed quantities (see quote, p. 191).

Definition in Terms of Limits

The standard modern approach to the definitions of the calculus replaces infinities and infinitesmals by limits that increase or decrease without bound. Aristotle had argued that in nature nothing can be actually infinite, but that the potentially infinite, "the open possibility of taking something more," was possible (*Phys.* 3.207a7–8). He gave as examples the number series 1, 2, 3,... and the unlimited ability to divide the line. Also, the Method of Exhaustion was based on the replacement of an actual infinity by a potential infinity.

[49] The "infidel mathematician" at whom Berkeley's diatribe was directed was Sir Edmond Halley (c. 1656–1743), the Astronomer Royal, after whom the comet is named, and the one most responsible for the publication of Newton's *Principia*: he encouraged Newton to write it and paid all the expenses. He also cataloged the stars and discussed their motion, and invented a diving bell. At the age of 22 he was made a Fellow of the Royal Society, but those hostile to his materialistic beliefs prevented him from being made the Savilian professor of astronomy at Oxford in 1691. Berkeley resented Halley for having convinced one of Berkeley's friends of the "inconceivability of the doctrines of Christianity" (Newman, *WM*, Vol. 1, p. 286).

Similarly, in the thirteenth century, Peter of Spain (1225–1277) distinguished *categorematic* (predicating) infinities from *syncategorematic* (jointly predicating) infinities (Clagett, RS); that is, a categorematic infinity is infinite in itself, whereas a syncategorematic infinity stands for a quantity larger (or smaller) than any other you choose. This of course is the infinity used in the modern limit concept. Swineshead, in his *Calculations*, had a similarly advanced notion of the "infinitely small" (*infinitum modicum*): rather than an actual infinitesmal it was an indefinite quantity smaller than any chosen. This is also in essence the interpretation that Kepler attached to his indivisibles. Other mathematicians, including John Wallis (1616–1703) and James Gregory (1638–1675) had argued for the use of limits, but were largely ignored. Yet others, including Colin Maclauren (1698–1746) and Joseph-Louis Lagrange (1736–1813), tried — unsuccessfully — to do without limits.

Newton's Use of Limits

Newton had a clear idea of infinitesmals as limits, for shortly after the previously quoted use of the "ultimate ratio of evanescent quantities" he explains:

> For those ultimate ratios with which quantities vanish are not truly the ratios of ultimate quantities, but limits towards which the ratios of quantities decreasing without limit do always converge; and to which they approach nearer than by any given difference . . . .

In explaining Newton's mathematics in the late eighteenth century D'Alembert (1717–1783) said, "The theory of limits is the true metaphysics of the calculus" (Kline, *MT*, p. 433). In retrospect it seems that Berkeley's criticism was unjustified, and we can sympathize with Newton's reluctance to publish for fear of "being bated by little smatterers in mathematics."

Nineteenth Century

In 1817 Bernhard Bolzano (1781–1848) gave the modern definition of the derivative:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

and in 1823 Cauchy gave the same definition. Yet there were still many logical problems in analysis because there was no agreed upon definition of a function, or even of a real number. These issues were addressed in the nineteenth century by Cauchy, Riemann, Gauss and other mathematicians, and finally assembled into an acceptable theory by Wierstrass in 1872.[50] For

[50] Augustin-Louis Cauchy: 1789–1857; Georg Friedrich Bernhard Riemann: 1826–1866; Carl Friedrich Gauss: 1777–1855; Karl Weierstrass: 1815–1897.

the time being, however, the infinitesmals and infinities had been banished (but see Section 5.3.6).

Manipulation of Infinitesmals

Kline (*MT*, p. 433) observes that eighteenth-century mathematicians did not clearly distinguish a very large number and an infinite number. For example, they treated infinite series just like finite series, with little concern for convergence. They were similarly cavalier in their use of infinitesmals, treating them as nonzeros when convenient (e.g. when dividing by them), and discarding them like zeros when that was convenient. A typical derivation of the derivative y' of $y = x^2$ might go like this. To find out how much x^2 changes relative to an infinitesmal increase o in x, write:

$$y' = \frac{(x+o)^2 - x^2}{o} = \frac{2xo + o^2}{o}.$$

Since o is infinitesmal but nonzero, we can do the division and write $y' = 2x + o$. Now since o is infinitesmal with respect to x we can discard it and conclude that the derivative of x^2 is $2x$, which it is. The modern derivation is not so different, but instead of treating it as neither zero nor nonzero, we take the limit as o goes to zero, and of course have to show that the limit exists.

Convenience of Infinitesmals

Although eighteenth- and nineteenth-century mathematicians were unable to come up with a consistent definition of infinitesmals and had to fall back on the use of limits, it was nevertheless true then, as it is now, that mathematicians and scientists think in terms of infinitesmals. It is very convenient to think in terms of the effect of varying some quantity x by an infinitesmal amount $\mathrm{d}x$. This is the reason that to this day we use Leibniz' notation for the calculus rather than Newton's.

Newton's Notation

Newton and Leibniz invented the infinitesmal calculus independently and almost simultaneously, which is no surprise since many of the ideas had been "in the air" for some years.[51] Of course, they devised different notations. Newton used dots to indicate derivatives, $\dot{x}$ for the first derivative of x, $\ddot{x}$ for the second derivative, etc. This notation is still occasionally used, especially in physics. For antiderivatives (integrals) he used dashes, $\acute{x}$ for the first antiderivative, $\acute{\acute{x}}$ for the second antiderivative, etc., a notation which is never seen now. Apparently Newton spent very little time worrying about notation; he wanted to get on with the physics.

Leibniz' Notation

[51] Newton seems to have done the work a little earlier (c. 1666 vs. c. 1673), though Leibniz published his first (1684 vs. 1687).

Leibniz, on the other hand, often spent weeks selecting a notation. We would expect as much, since we've seen that from his youth he had been designing calculi to facilitate reasoning (Section 4.3). He applied this experience well in the infinitesmal calculus. For example, the notation for a differential $\mathrm{d}x$ technically has meaning only as part of a derivative or an integral, $\mathrm{d}y/\mathrm{d}x$ or $\int y\,\mathrm{d}x$, but with a little care we can manipulate it as though it stands for a quantity. For example, we can "multiply" by it:

$$\frac{\mathrm{d}z}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}z}{\mathrm{d}x}, \qquad \frac{\mathrm{d}(u+v)}{\mathrm{d}x} = \frac{\mathrm{d}u}{\mathrm{d}x} + \frac{\mathrm{d}v}{\mathrm{d}x}.$$

This is especially important because the quantity $\mathrm{d}x$ mimics an infinitesmal, and so mathematicians can do correct calculus without abandoning their habit of thinking in terms of infinitesmals. The notation supports a useful, though technically fallacious, reasoning process.[52]

Formalism

The observation is important and bears repeating: In standard mathematics the notation $\mathrm{d}x$ has no meaning in isolation. Nevertheless it can be manipulated (according to the rules of the calculus) as though it has meaning, and still lead to correct results. This shows the power of formalism. Although $\mathrm{d}x$ has no *denotative meaning* — it does not stand for any quantity — it gets a *formal* meaning through its use in formulas according to the rules of the calculus. But if this works for $\mathrm{d}x$, why not for other bothersome mathematical objects? Why worry whether there really is a number $i = \sqrt{-1}$ so long as we can define rules for manipulating the symbol consistently? Indeed, why should we care whether *any* of the symbols of mathematics stand for anything, so long as they follow the appropriate rules of calculation? This is an important philosophy of mathematics, called *formalism*, that we'll consider later (Section 6.1.4).

The Priority Dispute

It is one of the sad events in the history of mathematics that the invention of the infinitesmal calculus was marred by a bitter priority dispute between its co-inventors — two of the most brilliant thinkers of their age; it's even more regrettable that the argument may have been instigated by third parties. However, the dispute had some interesting consequences that illustrate the forces that affect the evolution of science.

Retardation of English Mathematics

First is the role of nationalism: mathematicians took sides in the dispute, not on the basis of the facts, but on the basis of their nationality; for the most

[52] We may be reminded of Peter of Spain's distinction. Although $\mathrm{d}x$ is syncategorematic, since it has meaning only in the context of derivatives or integrals, it behaves in most cases as though it is categorematic, that is, as though its meaning is context-free.

part English mathematicians sided with Newton and continental mathematicians with Leibniz. (One wonders why most mathematicians even cared, but that they did is another fact worth consideration.) This "loyalty" extended even to the notation, with English mathematicians using Newton's dots and dashes, and continental mathematicians using Leibniz' much superior differential notation. As a result, "not only did the English mathematicians fall behind, but mathematics was deprived of contributions that some of the ablest minds might have made" (Kline, *MT*, p. 380–381). Such are the forces governing the progress of mathematics.

5.3.3 The Rational Numbers

History

It is a curious fact (the implications of which we'll consider later) that the development of rigor in mathematics proceeded from the apparently more complicated to the apparently simpler: first the calculus, then the irrational numbers, then the rational numbers, and finally the integers.[53] Nevertheless, I will deviate from the historical order, and discuss the construction of the rationals before that of the irrationals, since the former will clarify the latter.

Why was it necessary to construct the rational numbers at all? Recall that in Euclidean mathematics a ratio was not a number; a ratio such as $m : n$ could appear in a proportion, such as $m : n \; :: \; m' : n'$, and these proportions could be manipulated in various ways; for example, they could test this proportion by using natural-number arithmetic to check $mn' = nm'$.[54] However, the ratio itself was thought of more as a *shape* than a *quantity* (recall the Pythagorean representation of ratios by rectangular figures, p. 31). As we've seen, decimal notation helped mathematicians to become more comfortable with rational numbers, since they were able to calculate with them reliably. Nevertheless, the problems with irrational numbers and lingering doubts about negative and imaginary numbers led a few mathematicians to look for a rigorous basis for the rational numbers. The first attempt was made by Martin Ohm in 1822, but we will consider the modern construction developed in the 1860s by Karl Weierstrass (1815–1897) when he was a highschool teacher.

Pythagorean-style Rational Arithmetic

To understand Weierstrass's construction, consider again the Pythagorean representation of the ratio $m : n$ by a rectangular figure with sides in that

[53]Sources for this section include Dedekind (*ETN*, pp. 1–27) and Kline (*MT*, pp. 987–989).

[54]By a *natural number* I mean a nonnegative integer, that is, 0, 1, 2, 3, ...

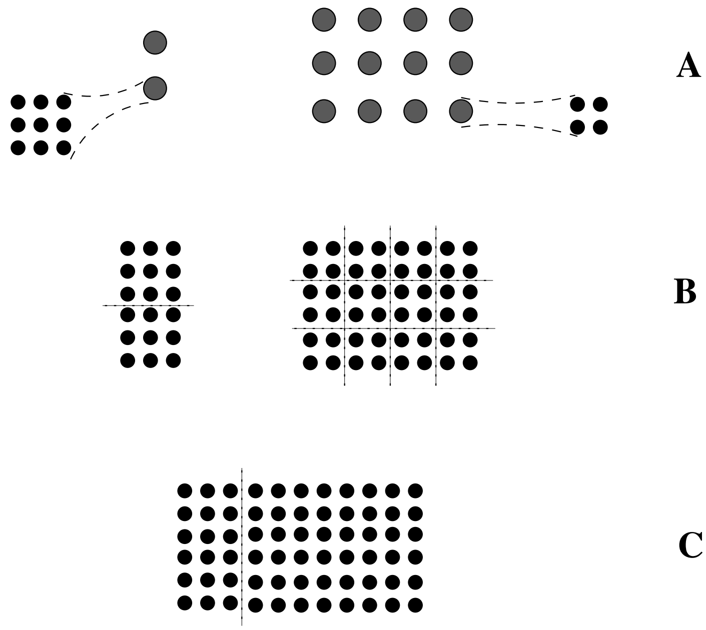


Figure 5.9: Example of addition of rational numbers represented by Pythagorean figures. **(A)** The figures to be added, $1 : 2$ and $4 : 3$. To add them they must be converted to a common denominator, so replace each term (pebble) in $1 : 2$ by a 3×3 square, and each term in $4 : 3$ by a 2×2 square. **(B)** The resulting figures $3 : 6$ and $8 : 6$, have the same height. **(C)** Combine the two figures to get the sum $11 : 6$.

ratio; that is, any km by kn rectangle represents the same ratio. Using these figures it's easy to define rules of calculation by which the pebbles (*calculi*) can be manipulated to implement rational arithmetic. For example, to multiply the ratio $p = m : n$ by the ratio $q = m' : n'$, replace each pebble in figure p by a copy of figure q. In modern notation,

$$pq = \frac{m}{n} \cdot \frac{m'}{n'} = \frac{mm'}{nn'}.$$

Similarly, to add p and q, we replace each pebble in p by a n' by n' square of pebbles, and each pebble in q by an $n \times n$ square, and then push the two figures together (Fig. 5.9). This is the familiar rule for adding fractions:

$$p + q = \frac{m}{n} + \frac{m'}{n'} = \frac{mn'}{nn'} + \frac{nm'}{nn'} = \frac{mn' + nm'}{nn'}.$$

This is *calculation* (pebble manipulation) in the most literal sense. (Exercise for the reader: define division p/q and subtraction $p - q$ in the Pythagorean way. What additional complications do you encounter?)

The essential point of these rules for calculation is that we can operate on a ratio $m : n$ by manipulating the natural numbers m and n. In effect the rational number m/n is functionally equivalent to the pair of natural numbers (m, n), given some peculiar rules for doing arithmetic on the pairs. This is the approach Weierstrass used.

Existence of Rational Numbers

In discussing Weierstrass's construction I will write rational numbers in the form '$m : n$' rather than the familiar m/n or $\frac{m}{n}$, since the latter suggest that we are dividing m by n. Whether we can divide one number by another that doesn't factor the first depends, of course, on the existence of rational numbers, so we must not beg the question by assuming the division makes sense. Thus we assume the existence only of the natural numbers and then *construct* the rational numbers from pairs of naturals.

Problem of Unique Representation

The essence of Weierstrass's construction is to represent positive rational numbers by pairs of natural numbers in the appropriate ratio.[55] Here we encounter the first difficulty, for many pairs of numbers represent the same rational number, for example,

$$1:2 \;=\; 2:4 \;=\; 3:6 \;=\; 4:8 \;=\; \ldots$$

There are many more pairs $m : n$ than there are rational numbers. This is the *problem of unique representation* and it arises whenever one formal structure is used to represent another; it's an important practical problem in computer science (MacLennan, *FP*, Ch. 4). In this case the problem is solved simply by saying that rational numbers correspond to pairs of integers *in lowest common terms*; in other words, whenever we calculate a rational number we always reduce it to lowest terms, just as we learned to do in grade school.[56]

Arithmetic

Given the correspondence between rational numbers and pairs of integers, rational arithmetic can be defined in terms of operations on natural numbers:

$$m : n \;+\; m' : n' \;\;=\;\; \mathsf{reduce}\{(mn' + nm') : nn'\}$$

[55]Negative rational numbers are considered later.

[56]Whole numbers are represented by ratios of the form $m : 1$. The mathematically more sophisticated solution is to define a rational number to be an *equivalence class* of pairs of integers, all the pairs in such a class being in the same ratio. This definition requires set theory, which hadn't been invented in the 1860s when Weierstrass did this work.

$$\begin{aligned} m:n \;-\; m':n' &= \mathsf{reduce}\{(mn'-nm'):nn'\} \\ m:n \;\times\; m':n' &= \mathsf{reduce}\{mm':nn'\} \\ m:n \;\div\; m':n' &= \mathsf{reduce}\{mn':nm'\} \end{aligned}$$

where by $\mathsf{reduce}\{m:n\}$ I mean that the pair $m:n$ must be reduced to lowest common terms.

Comparison

It is also easy to compare rational numbers in terms of their natural number constituents. For example, $m:n \;=\; m':n'$ if and only if $mn' = nm'$, and $m:n \;<\; m':n'$ if and only if $mn' < nm'$. The latter comparison is important, since we have defined only nonnegative rational numbers, and so $p-q$ makes sense only if $p \geq q$.

Negative Rationals

Some nineteenth century mathematicians were still uncomfortable with negative numbers, and we ourselves would have trouble calculating with negative numbers in the Pythagorean way. We can avoid these difficulties and illustrate a purely formal solution to the problem by simply stipulating that negative numbers are the same as positive, but differently "colored." If we were doing Pythagorean calculation, we could literally make the "negative pebbles" a different color. They are no less there than the "positive pebbles," they simply obey different rules; adding three negative pebbles is the same as taking away three positive ones, etc.

Since we no longer do mathematics concretely (!!) like the Pythagoreans, we can say that a negative rational number is a pair of natural numbers with some additional mark attached; I will write $\overline{m:n}$. It is then straight-forward (but tedious) to define rational arithmetic on all possible combinations of positives and negatives:

$$\begin{aligned} \overline{m:n} \;\times\; m':n' &= \overline{mm':nn'} \\ \overline{m:n} \;\times\; \overline{m':n'} &= mm':nn' \\ &\text{etc.} \end{aligned}$$

Correctness of Construction

Given this construction of the rational numbers, it's straight-forward to show that they obey all the expected laws of rational numbers (commutative, associative, distributive, etc.) *assuming* the familiar properties of natural numbers; it's an entertaining exercise, but unnecessary for our purposes.[57]

Dense but not Continuous

[57]In programming language terms, we have implemented nonnegative rationals as the *direct product* of the naturals with themselves (i.e., as records with two natural number

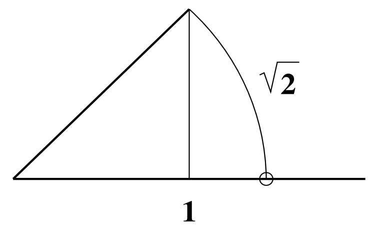


Figure 5.10: Although the rationals are dense, they do not include all the points on the line. As the Pythagoreans proved, the point corresponding to $\sqrt{2}$ does not correspond to any rational number (where the leg of the triangle corresponds to the unit). Nevertheless, there is an infinity of rationals between any two rationals.

Since the Weierstrass construction allows us to add any two rational numbers and to divide any rational number by two, we see that between any two rationals p, q there is a rational $\frac{p+q}{2}$. In brief, between any two rationals there is always another. In fact, since this process of division can be continued as long as we please, between any two rationals there is an *infinity* of rational numbers. But such infinite divisibility — now called *denseness* — had often been taken as the hallmark of the continuum, an opinion expressed as late as Bolzano (1781–1848). There is more to continuity than this, since although the rationals are *dense*, they do not include all the points on the line (as the Pythagoreans knew, Fig. 5.10). In other words, the rational numbers are *incomplete* because, although they are dense, there are still "gaps." How then *can* we ensure that we have captured all the points of the continuum?

5.3.4 Plato: The Monad and the Dyad

> Universal nature molds form and type by the constant revolution of potency and its converse about the double in the various progressions.

fields), and have implemented the full rationals as the *direct sum* (discriminated union) of the nonnegative rationals with themselves. These are standard computer science techniques, and indeed, the implementation of rational number arithmetic is a common programming exercise. The issues we've touched on here, such as the problem of unique representation and proving the correctness of an implementation, are typical of those encountered in defining *abstract data types* in computer science (MacLennan, *FP*, Chs. 2, 4, 5).

— Plato, *Epinomis* 990e

> I mean by 'continuous' that which can be divided into things that can be divided forever.
>
> — Aristotle, *Physics* 6.2.232b25-26

Many Definitions of Irrationals

In the period 1858–1872 a number of mathematicians, including Weierstrass, Dedekind, Méray, Heine and Cantor, offered definitions of continuity, which are essentially equivalent.[58] They all built rather directly on Euclid, and most of the solutions could have been discovered by Greek mathematicians, had they fewer philosophical scruples (we'll see below how close Plato came). I will present Cantor's solution, since it will prepare the way for other issues that we must consider.

Plato's Goal

Before we investigate Cantor's construction of the irrational numbers, it will be enlightening to look at Plato's approach to the problem.[59] We saw in Chapter 2 how the Pythagorean discovery of incommensurable magnitudes led Aristotle to separately axiomatize discrete and continuous quantity, that is, numbers (*arithmoi*) and magnitudes (*megathoi*). This decision may have been influenced by an unsuccessful Platonic attempt to find a single set of first principles (*archai*) for both rational and irrational numbers, in effect, the first attempt to construct the complete real number system.

Monad and Dyad

Recall that the Pythagoreans saw a fundamental opposition in nature, between the (de)finite and the in(de)finite, and that these were the first principles of all things. In the more mathematical Platonic brand of Pythagoreanism these two principles became the *monad* ($\mu o\nu \acute{\alpha} \varsigma$) and the *dyad* ($\delta \upsilon \acute{\alpha} \varsigma$). The monad is the fundamental principle of *identity*, and so is the delimiting, form-giving principle. In contrast, the dyad is the fundamental principle of *nonidentity* or *difference*, and so is the undelimited, form-receiving princi-

[58] Richard Dedekind: 1831–1916, Charles Méray: 1835–1911, (Heinrich) Eduard Heine: 1821–1881, Georg Cantor: 1845–1918.

[59] Much of this account is conjectural and the work of A. E. Taylor (*PS*, Ch. 3), summarized in Maziarz & Greenwood (*GMP*, Ch. 14). Taylor is building on work by Burnet, Milhaud and Stenzel. The primary basis for his argument is the *Epinomis* (990c5–991b4), a dialog that many scholars think was not written by Plato, but which Taylor accepts as authentic. In any case, it's certainly Platonic in content.

ple.[60] Further, the monad is the principle of stability or *being*, and the dyad the principle of change or *becoming*. This is illustrated by the square and rectangular numbers (Figs. 2.4, 2.5, p. 23), since the "rule" brings forth stability from the monad, but change from the dyad. Plato thought that from these two principles he could construct all the other forms, as well as explain the way in which the forms (monad) are imposed on matter (dyad). Here we will only be considered with his attempt to generate the real numbers from the monad and the dyad.

The Dyad as an Operator

The dyad under consideration here is the *indeterminate dyad*, which we must carefully distinguish from the absolute number 2, or *auto-dyad*.[61] The indeterminate dyad is not a definite number, but rather an operator for generating numbers; it is the basic principle of change. In some contexts it is called "two-making" (*duopoios*) and can be applied in different ways to change something in a twofold way; for example it can be used to double a number, or to halve it, or to split the difference between two numbers (i.e. take their average or arithmetic mean).[62] These ideas will be illustrated by showing how the monad and dyad can be used to construct the natural numbers. The indeterminate dyad does not represent any definite number, but it can be made determinate by applying it to the only number we have to start with, 1, represented by the monad. Thus, if we let dn represent using the dyad to double n, then $d1$ is the determinate number 2. More carefully, since d is the principle of change, so $d1$ is a different number from 1, which

[60] As noted previously (p.31, n. 14) the two Pythagorean principles, definite and indefinite, correspond closely to two fundamental forces of Chinese philosophy, yang and yin. In this connection it's interesting to note that yang, the creative principle — corresponding to the Pythagorean Monad — is represented in the *I Ching* by an unbroken line: ━━, whereas yin, the receptive principle — corresponding to the Dyad — is represented by a broken line: ━ ━. Philosophical use of yang and yin begins in the fourth century BCE, whereas the Pythagorean doctrines date from the sixth century BCE; on the other hand the text of the *I Ching* may go back to the sixth century BCE (Ronan, *Science*, p. 145; Schwartz, *WTAC*, pp. 40, 180–181, 187, 203, 390–400, 440 n. 1). It's unlikely however that either influenced the other; more likely they come from common intuitions about the world (Burkert, *LSAP*, pp. 472–474).

[61] The indeterminate dyad is *ἀόριστος δυάς* and the auto-dyad is *αὐτὸ ὅ ἐστι δυάς*. See also Burkert (*LSAP*, pp. 35–37).

[62] These ideas are still current computer science in *functional programming*, where for example we must distinguish the number 2 from the constant 2 function and from the doubling, having and averaging functions (MacLennan, *FP*, pp. 240–242, 262–265, 276–277).

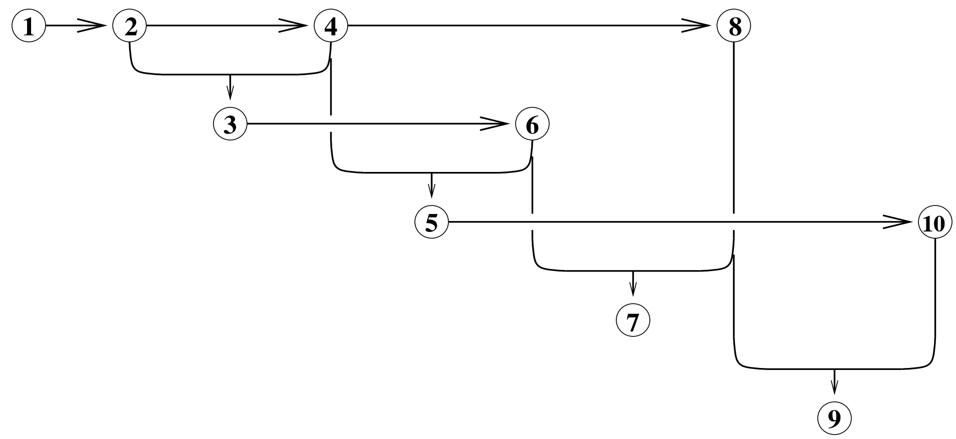


Figure 5.11: Generation of Natural Numbers from Monad and Dyad. Horizontal lines indicate use of the indeterminate dyad to produce the double of a number; vertical lines indicate the use of the indeterminate dyad to produce the average of two numbers. By these two operations all the natural numbers can be generated from the monad (1).

we may choose to call 2.[63] A second application of the dyad, $dd1$, produces a number that can be called 4. Using the dyad in a different way we average 2 and 4 to get 3, which we write $3 = D\,2\,4$. Continuing in this way we can define all the natural numbers, though they are not generated in consecutive order (Fig. 5.11):

$$\begin{array}{rcccl}
1 & = & 1 & = & 1 \\
2 & = & d1 & = & d1 \\
4 & = & d2 & = & dd1 \\
3 & = & D\,2\,4 & = & D\,d1\,dd1 \\
6 & = & d3 & = & d\,Dd1dd1 \\
8 & = & d4 & = & ddd1 \\
7 & = & D68 & = & D\,dDd1dd1\,ddd1 \\
 & \vdots & & \vdots &
\end{array}$$

This is a novel way of generating the natural numbers, since it uses an essentially multiplicative process rather than the usual additive process ("add

[63]The monad is an ideal form, not an ensemble, so we cannot "add" the monad to itself to get 2.

one"). The dyad, in its general role as *other-making* (δ), could have been used to generate the natural numbers in succession, $2 = \delta 1, 3 = \delta 2 = \delta\delta 1$, etc. However, Plato wanted to generate *all* the real numbers, and for this he needed the multiplicative dyad.[64]

The Dyad as an Approximating Sequence

The dyad can also be used to generate the rational and irrational numbers. For example, we can use it to divide the monad thus generating 1/2; applying it again yields 1/4 and averaging 1/2 and 1 yields 3/4. The result is a kind of binary notation. We have seen (p. 32) that Theon of Smyrna defined an infinite sequence for $\sqrt{2}$, which may have been known to Plato. It is significant that this series is alternately above and below its limit, since Taylor thinks Plato took the series as a prototype for oscillating approximations to any real number. If we write such a series $(a_1, b_1, a_2, b_2, \ldots)$, where the rational approximations a_k are above the limit and the b_k are below it, then the dyads $[a_k, b_k]$ are progressively narrowers brackets around the limit. The approximation process can continue as long as there is a *gap* between the two, a dyad, which he calls a "great-and-small." If the sequence reaches its limit (i.e., it has a finite binary representation), then for some k the dyad $[a_k, b_k]$ collapses to a monad $a_k = b_k$. This is another example of how the monad limits the otherwise unlimited dyad. However, for irrational numbers (and certain rational numbers) there will always remain a "gap," a residual of becoming, matter, and dyad that has not been reduced to being, form, and monad.

Plato's Arithmetization of the Real Numbers

What has Plato accomplished? He has succeeded in generating the entire (positive) real number system — natural, rational and irrational numbers — from two principles, the monad and the dyad (though he uses the dyad in three different ways). Taylor (*PS*, p. 120) says,

> The Platonic theory is inspired by the same demand for pure rationality which has led in modern times to the "arithmetisation of mathematics". The object aimed at, in both cases, is to get rid of the dualism between so called "continuous" and "discrete" magnitude.

Plato's arithmetization of the real numbers was criticized by Aristotle (*Met.*

[64]The natural numbers are easily generated from the Pythagorean first principles, the definite and indefinite. As the undifferentiated background, the indefinite corresponds to 0, *no thing*. Applying the definite ($\mathcal{D}$) to the indefinite determines a thing, and thus creates $1 = \mathcal{D}0$. Applying the definite again makes a discrimination and further determines it, creating $2 = \mathcal{D}1 = \mathcal{D}\mathcal{D}0$, and so forth.

M 1080a13–1085a3) because it lacks a property Aristotle demands of any theory of the natural numbers: their sequential generation. Therefore Aristotle rejected Plato's attempted unification of numbers and magnitudes, and this decision, enshrined in Euclid, set the direction for mathematics down through the seventeenth century (Section 5.1).

5.3.5 Cantor: The Real Line

> I made the fixed resolve to keep meditating on the question till I should find a purely arithmetic and perfectly rigorous foundation for the principles of infinitesmal analysis.
>
> — Dedekind (*Essays on the Theory of Numbers*)

Irrationals as Sequences of Rationals

Like Plato, many mathematicians in the mid-nineteenth century tried to define the irrationals as the limits of convergent sequences of rational numbers.[65] However, Cantor pointed out the logical fallacy in this approach: you can only define an irrational to be the limit of a sequence of rationals is you know that the sequence has a limit; but the existence of irrational limits was just what was to be proved. These mathematicians had begged the question. Cantor avoided this problem of the existence of a limit by defining an irrational number as the convergent sequence *itself*. We'll look at this approach in some detail because it bears on the development of constructive analysis, which is closely related to the theory of computation.

Fundamental Sequences

Cantor defines a *fundamental sequence* X to be a sequence of rational numbers, $X = (p_1, p_2, p_3, \ldots)$, such that

$$\lim_{n\to\infty} p_{n+m} - p_n \quad = \quad 0, \qquad \text{for all } m.$$

That is, for any positive number ϵ we might choose, the elements of the sequence will eventually differ by less than ϵ (Fig. 5.12). For example, the sequence of rationals

$$\Theta = \left(\frac{1}{1}, \frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \frac{99}{70}, \frac{239}{169}, \ldots\right) \tag{5.1}$$

defined by Theon of Smyrna is a fundamental sequence.

[65] Sources for this section include Kline (*MT*, Ch. 41).

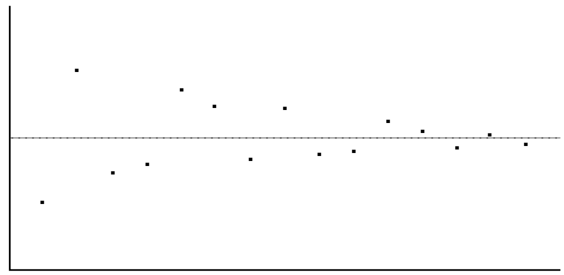


Figure 5.12: Example of a fundamental sequence, which is a sequence of rational numbers that will eventually differ by less than any positive number we choose. A fundamental sequence can be identified with the real number to which it converges.

A fundamental sequence X is *convergent* (in Cauchy's sense), and it is reasonable to assume that there is some real number x to which it converges, but this begs the question of the completeness of the number system. So instead Cantor identifies a fundamental sequence with the real number that we would like to be its limit. Roughly speaking, Cantor *defines* real numbers to be fundamental sequences. Thus, instead of *postulating* a number $\sqrt{2}$ as the limit of the sequence Θ, he *defines* $\sqrt{2}$ to be the sequence Θ. Of course, this is not enough in itself; it's also necessary to show that one can operate on fundamental sequences as on real numbers and that the set of all fundamental sequences are complete (as the rationals were not).

Identity and Equality

It's apparent of course that, since many fundamental sequences X, X', X'', ... correspond to each real number x, we can't in fact identify reals and fundamental sequences, since there are too many of the latter. For example, in addition to the sequence in Eq. 5.1, this one also represents $\sqrt{2}$:

$$\left(\frac{1}{1}, \frac{14}{10}, \frac{141}{100}, \frac{1414}{1000}, \frac{14\,142}{10\,000}, \frac{141\,421}{100\,000}, \ldots\right).$$

Cantor's solution is to identify a real number with the *equivalence class* of all fundamental sequences with the same limit. The latter notion must be defined without reference to the real number that is the limit, since Cantor says two fundamental sequences X and Y have the same limit if and only if

$$\lim_{n \to \infty} |X_n - Y_n| = 0.$$

This amounts to a definition of the equality of the real numbers x and y corresponding to X and Y. Comparison between reals $x < y$ is also defined in terms of their fundamental sequences; it is of course necessary to show that the result of a comparison does not depend on the particular sequences chosen as the representatives of the numbers. In fact, one can show that, just like the points on the line, and two fundamental sequences X and Y must satisfy exactly one of the relations $X = Y$, $X < Y$ or $Y < X$.

Arithmetic

Just as in the construction of the rationals, arithmetic operations were defined in terms of arithmetic on the paris of natural numbers representing the rationals, so in Cantor's construction, arithmetic on real numbers is defined in terms of arithmetic on fundamental sequences that represent the reals. For example, if the fundamental sequences X and Y represent the reals x and y, then the sum $x + y$ is represented by

$$(X_1 + Y_1,\ X_2 + Y_2,\ X_3 + Y_3,\ \ldots),$$

which can be shown to be a fundamental sequence. Similarly the product xy is represented by the fundamental sequence $X_n Y_n$. All the arithmetic operations can be defined in this way, and can be shown to obey the laws of real arithmetic (commutative, associative, distributive, etc.)

Completeness

We saw that the rational number system is incomplete, that it has "gaps" in it. In particular it was missing the limits of some convergent sequences of rationals, namely those whose limits are irrational, such as $\sqrt{2}$. With Cantor's construction, these limits are included (by definition), but how can we be sure there aren't other "gaps"? For example, we started with the rationals, and by taking Cauchy-convergent sequences of these we got some more numbers, which we added to the number system. Perhaps if we took Cauchy-convergent sequences of these new numbers, we would get a yet newer set, which would also have to be added. A sequence of these might yield additional numbers, and so forth without end.

Cantor proves that this circumstance doesn't arise by showing that the limits of Cauchy-convergent sequences of *real* numbers are already in the number system. More precisely, he shows that there is a fundamental sequence (of *rationals*) corresponding to the limit of any Cauchy-convergent sequence of *real* numbers, and that the rational elements of the limit can be constructed from the rational elements of the members of the reals in the sequence. Thus we do not have to worry about a real sequence converging to a limit that's missing; there are no gaps, it is *complete*.

5.3.6 Infinities and Infinitesmals

> Perhaps indeed, enthusiasm for nonstandard methods is not unrelated to the well-known pleasures of the illicit.
>
> — Martin Davis

Problem of Infinite Representation

The arithmetization of geometry was accomplished by Cantor, Dedekind, Weierstrass and others, but only at the expense of using infinite objects. For example, in Cantor's approach real numbers are represented by infinite sequences; in Dedekind's they are represented by pairs of infinite sets (so called "Dedekind cuts"). In this sense these mathematicians had not solved the problem discovered by the Pythagoreans, namely that no *finite* figure could represent $\sqrt{2}$.

Potential Infinity

Certainly, some infinite sequences are only *potentially infinite*, that is, they are an unending process resulting from a finitely specifiable rule. For example, Theon's sequence Θ is given by the rules:

(1) $\Theta_1 = \frac{1}{1}$;

(2) if $\Theta_k = \frac{m}{n}$ then $\Theta_{k+1} = \frac{m+2n}{m+n}$.

Such sequences are no more problematic than the infinite sequence of natural numbers, which is also only potentially infinite, since it is generated by the simple rule, "add one."

Uncomputable Reals

The use of infinite sequences raises two questions. First, is there any guarantee that all sequences can be described by a finite rule? Or, more to the point, is there any guarantee that each real number is represented by at least one such sequence? In fact, the answer to both questions is "no," but mathematicians at that time did not know so (although Cantor laid the foundation for these results, which are presented in the next chapter). Any sequence that cannot be described by a finite rule is an *actual infinity* rather than a potential infinity, and so there isn't a finite representation for any reals requiring such sequences for their representation, and the Pythagorean problem remains for these.

What is a Rule?

The second problem was that the idea of a "rule" was itself informal, so mathematicians were unable to say precisely what it means for a sequence to be finite specifiable. The attempt to formalize these ideas ultimately led in the early twentieth century to the axiomatization of mathematics and theory of computability, which is the foundation of computer science (Chapter 6). Most mathematicians were not especially bothered by these actual infinities; indeed, theories of infinite sets and infinite numbers became widely accepted tools in mathematics.

Objections to Arithmetization

On the other hand, not all mathematicians agreed with the reduction of the real continuum — based as it was on clear intuitions of time and geometry — to complicated constructions of the rationals and integers, and considered "the irrational number, logically defined, [to be] an intellectual monster" (Kline, *MT*, p. 987). For example, in 1887 the French mathematician E. H. Du Bois-Reymond (1818–1896) objected to the replacement of magnitudes by a real arithmetic constructed "with help from so-called axioms, from conventions, from philosophic propositions contributed *ad hoc*, from unintelligible extensions of originally clear concepts" (Kline, *MT*, p. 992), and in 1867 the German mathematician Hermann Hankel (1839–1873) said:

> Every attempt to treat the irrational numbers formally and without the concept of magnitude must lead to the most abstruse and troublesome artificialities, which even if they can be carried through with complete rigor, as we have every right to doubt, do not have a higher scientific value. (Kline, *MT*, p. 987)

In fact they do have scientific value, since they elucidate the relation of the continuous and the discrete. There was nevertheless some truth in these objections, but they were destined to be lost in the stampede to reduce mathematics to the discrete.[66]

Nonstandard Analysis

I mentioned previously that one of the advantages of Leibniz' differential notation was that in many cases it permits you to manipulate differentials as though they stand for infinitesmal quantities, and so to think in familiar ways about problems in the calculus. This might lead you to wonder if there is some way to fix up the rules of the calculus so that the use of differentials always yields correct results; this would argue for the acceptance of infinitesmals as legitimate mathematical objects, at least from a formalist perspective.

Something very similar to this was accomplished in the early 1960s by Abraham Robinson (1918–1975) in his *nonstandard analysis*.[67] Using the

[66] Even Hilbert proposed a direct axiomatization of the reals that did not reduce them to the rationals; see his article in the *Jahres. der Deut. Math.-Verein.*, 8 (1899), pp. 180–184, reprinted in his *Grundlagen der Geometrie*, 7th ed., App. 6.

[67] Robinson (*NSA*) is the primary source for nonstandard analysis; a readable overview

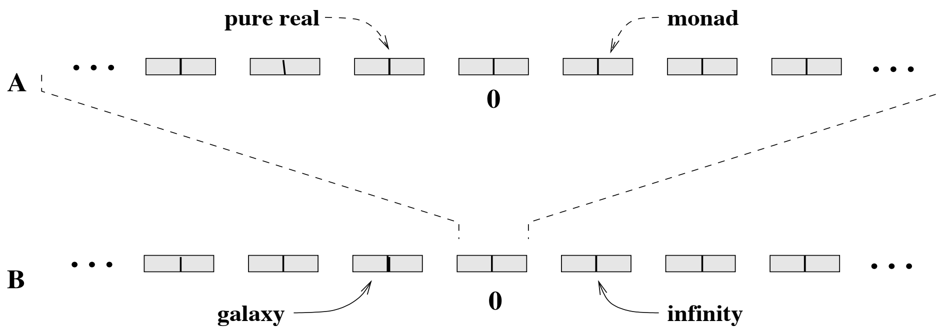


Figure 5.13: Visualization of Hyperreal Number System. **(A)** Each pure (standard) real number x (indicated by thick vertical lines) is surrounded by a monad, a cluster of infinitesmally close hyperreal numbers of the form $x + \mathrm{d}x$. **(B)** The reciprocals of infinitesmals are infinities $\mathrm{d}x^{-1}$ (indicated by thick vertical lines), each of which is surrounded by a galaxy, a cluster of hyperreal numbers of the form $\mathrm{d}x^{-1} + x + \mathrm{d}x$.

framework of mathematical logic, and important theorems proved about it in the early twentieth century (discussed in Ch. 7), Robinson showed how to construct from the reals a system of numbers containing infinitesmals, just as Cantor had constructed the reals from the rationals, and Wierstrass the rationals from the integers. The exact method is not relevant to our purposes here; suffice it to say that the result is a set of infinitesmal quantities $\mathrm{d}x$ that are nonzero yet smaller than any real number y, $0 < |\mathrm{d}x| < |y|$. Thus each "pure" real number y is surrounded by a cluster of infinitely close *hyperreal numbers* of the form $y + \mathrm{d}x$. Robinson calls these clusters *monads*, following Leibniz. Although we think of the real line comprising contiguous real numbers, in nonstandard analysis we must picture the pure real numbers as surrounded by nonoverlapping monads (Fig. 5.13A). Furthermore, for each infinitesmal $\mathrm{d}z$ there is a corresponding infinity, its reciprocal, which we may write $\mathrm{d}z^{-1}$. Since these infinities are infinitely far apart, each has an associated set of hyperreal numbers of the form $\mathrm{d}z^{-1} + y + \mathrm{d}x$, for any hyperreal $y + \mathrm{d}x$; these clusters are called *galaxies* (Fig. 5.13B).

Nonstandard Elementary Calculus Texts

Nonstandard analysis legitimates reasoning in terms of infinitesmals and infinities, which has proven to be an intuitive way of thinking about the

is Davis & Hersh (*ME*, pp. 237–254). See also Davis (*ANSA*), Hurd & Loeb (*INSRA*), Lutz (*NSA*) and Stroyan & Luxemburg (*ITI*).

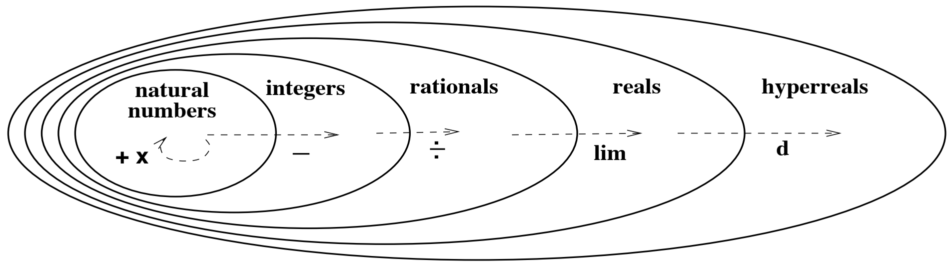


Figure 5.14: Relative Completeness of Number Systems. The natural numbers are complete (closed) under addition and multiplication, but not subtraction. The integers are complete under all three operations, but not under division, which leads to the rational numbers. Similarly, the standard reals are complete with respect to convergent sequences, but not with respect to operations that produce infinitesmals.

continuum, in spite of attacks through the ages by Zeno, Berkeley and others. For this reason several elementary calculus books have appeared that explain the derivative and integral by means of infinitesmals, as did the masters of the seventeenth century, rather than by means of the traditional, unintuitive limits.[68]

Completeness of Reals

You may be perplexed by the picture of the continuum provided by nonstandard analysis. I previously claimed that Cantor's construction filled in all the "gaps" between the rational numbers, and that Cantor proved that the resulting real number system was "complete." Yet from the standpoint of nonstandard analysis there are gaps in the real line, for it is in these gaps that we find the infinitesmals. Does this mean that Cantor's proof was wrong?

Cantor proved that the (standard) reals are complete in a very specific sense: under the formation of convergent sequences; he showed that no further gaps would be revealed by sequences of reals. Thus, the idea of "completeness" is relative to a certain class of operations; a set is complete if it is *closed* under these operations, that is, they cannot take us out of the set. For example, even the rationals are in a sense complete, for they are closed under finite arithmetic operations: only rational numbers will result from adding, subtracting, multiplying or dividing rational numbers. The integers are closed under addition, subtraction and multiplication (Fig. 5.14). What nonstandard analysis reveals is that there are other operations, such as the

[68]See, for example, Keisler (*EC*) and the instructor's guide, Keisler (*FIC*).

differential, that take us out of the standard reals and into the hyperreals. All of which should leave us wondering if our discrete symbols have yet captured all that the continuum has to offer.[69]

5.4 Summary

5.4.1 Technical Progress

Progress on the Pythagorean Program

The sixteenth and seventeenth centuries made significant progress in two goals of the Pythagorean research program: the reduction of the continuous to the discrete, and the mathematization of nature. After more than twenty centuries arithmetic and geometry were reunited. Geometry had been reduced to algebra, and the continuum was purged of its indivisibles. A technical device, the limit, allowed irrationals to be reduced to a potential infinity of rationals, which were in turn reduced to integers. Likewise, the first steps were made in the mathematization of nature, in Plato's terms, making the world of Becoming intelligible by explanation in terms of the real of Being. This was the explanation of motion in terms of ratios, the first successful quantification — reduction to calculation — of change.

Algebra and Formalism

The development of algebra showed the value of variables, which still play an important part in most rule based systems, since they allow a rule to be general, that is applicable to many individuals or data. The use of algebraic notation to describe indifferently the manipulation of rational and irrational numbers and magnitudes, and the use of infinitesimal notation in the calculus, showed that in many cases *notation can do work* merely by virtue of its syntax, or form; it is not necessary that the symbols have a meaning; calculation is independent of semantics. Thus the power of formality was made manifest, which paved the way for the development of symbolic logic by Boole and others. Algebraic techniques demonstrated how pure mechanical symbol manipulation could facilitate and even mimic the operations of thought.

Algebra also showed how mathematical symbols could be given meaning by their algebraic or formal properties, which side-stepped many of the philosophical difficulties. This led to *formalism*, an important philosophy of

[69] Hamming (UEM) has a nice discussion of how nonclosure under various operations had led to the progressive extension of the number system.

mathematics, which motivated the development of the theory of computation; it is discussed in the next chapter.

5.4.2 Psychology and Sociology of Science

Epistemology: Structure and Process

The subject of this book is epistemology — the theory of knowledge, not just the static structure of knowledge, but also the dynamic processes by which it evolves. In this chapter we have seen examples of these processes from the real history of science, as opposed to the cartoon version that is often presented.

Role of Metaphysical Commitment

We have seen the role of metaphysical commitment — the belief that a theory describes the actual nature of reality — and its absence, as advocated by positivism. Positivism may be liberating or stultifying. Sometimes a positivist attitude — a suspension of belief or disbelief, "putting aside the mental torments" — allows one to bypass metaphysical objections and make progress. Because the Arabs were unmoved by Greek philosophical worries, they were able to begin treating irrationals as numbers; because sixteenth-century mathematicians trusted their symbols, negative and imaginary numbers came to be accepted; because seventeenth-century mathematicians naively used infinitesimals, they made progress in the calculus and its application to nature.

On the other hand, simply "saving the phenomena" may not create the emotional commitment needed for major scientific advances. The scientific pioneers all had comprehensive beliefs about the nature of the universe, and believed that they were discovering the secrets of the cosmos or even looking into the "mind of God." It doesn't much matter whether these beliefs were true or false (in our view), they still had their effects. Neoplatonic mysticism encouraged the search for mathematical laws; a belief in the harmony of the spheres led to Kepler's Laws; Hermetic reverence for the sun opened the way for heliocentrism; alchemy encouraged experiment and applied science; Newton's acceptance of "subtle spirits" pervading the universe helped him postulate gravity. Although positivist *methods* have sometimes led to progress (e.g., in Einstein's development of special and general relativity), the major figures in *science* (including Einstein), *believed* their theories. *No great scientist was a positivist.*

Other Psychological and Social Factors

We have also seen how progress can be impeded by extraneous factors; for example, notational allegiances motivated by the priority dispute between Newton and Leibniz held back English mathematicians and impeded the progress of mathematics overall. On the other hand, progress can also be stimulated by events apparently antithetical to it, such as intolerance and the suppression of ideas. The condemnation of 1277 loosened the Aristotelian strangle-hold on science, and allowed new theories to be contemplated. Even the witch-hunts — though a foul blot on European history — encouraged science to separate itself from magic and religion, and to better define its methods.

Finally, the consequences of science are not always as expected. Newton thought that his investigations of occult forces would refute atheism forever, yet his theory of gravity and laws of motion established precisely the mechanistic and deterministic worldview that Berkeley feared as a threat to religion.

Chapter 6

Theory of Computation

6.1 Philosophies of Mathematics

> Whatever is provable ought not to be believed in science without proof.
>
> — Richard Dedekind, *Was sind und was sollen die Zahlen?* (1888)

> Mathematics, rightly viewed, possesses not only truth, but supreme beauty — a beauty cold and austere, like that of sculpture.
>
> — Bertrand Russell, *Mysticism and Logic* (1908)

6.1.1 Peano and the Move Toward Rigor

We have been tracing the parallel, but cross-linked developments of epistemology and mathematics over the 2500 years since Pythagoras first demonstrated the reduction of expertise to calculation. In this chapter we continue on the mathematical track by considering the formalization of mathematics, which motivated the development of the theory of computation and led to the discovery of the limitations of discrete symbol manipulation systems, which we'll take up in Chapter 7.[1]

The Loss of Rigor

[1]Sources for this section include Kneebone (*MLFM*, Chs. 5, 6).

Euclid had constructed arithmetic and geometry on separate foundations, and of the two, geometry's was more secure. This security had been undermined in the seventeenth century by the arithmetization of geometry and the spread of algebraic methods throughout mathematics. The result was one of mathematics' most fertile epochs, but by the end of the eighteenth century some mathematicians worried about the lost rigor. Through the first half of the nineteenth century Legendre, Lagrange, Gauss, Cauchy and Weierstrass all worked to strengthen the deductive structure of the new analysis, but they accepted arithmetic and algebra without question. Towards the end of the nineteenth century several mathematicians pushed deeper into the foundations: we have already seen Cantor's construction of the real numbers from the rationals, and Weierstrass's construction of the rationals from the natural numbers. In a process of intellectual development that set the pattern for mathematics as a field, Dedekind first secured the reals by constructing them from the rationals (1872), and then delved into the construction of the rationals and natural numbers (1888).[2]

Peano: 1858–1932

Giuseppe Peano was the catalyst for the movement to reduce mathematics to a calculus. He used more rigorous methods than his contemporaries in his 1884 book on the infinitesimal calculus, and he was skilled at constructing counterexamples to disprove claims based on intuitive reasoning.[3] Since intuitive reasoning was always vulnerable to the construction of counterintuitive

[2] The first in his *Stetigkeit und irrationale Zahlen* (Continuity and Irrational Numbers); the second in his *Was sind und was sollen die Zahlen?* (What are the Numbers and What is their Purpose?).

[3] One example is his famous space-filling curve. When such a "monster" is presented as a counterexample to a hypothesis, one response is to accommodate it, either by extending the theorem, or by placing additional conditions on the hypothesis so that the monster is excluded. Thus we might state the theorem as applying to all curves having such-and-such a property (which excludes the space-filling curve). Another response — called "monster-barring" — is to claim that such a thing is not a curve at all, so of course it's not a counterexample. The first response could be called "liberal," in that it accepts the opponent's claim that the monster is a member of the class in question, and then tries to accommodate the monster by extending the theorem or by placing additional conditions on it. The second response can be called "conservative," since in effect it says, "Never before have we considered such monsters to be members of this class, nor are we going to now." The class can then be defined in such a way that the monster is excluded. In fact, it is most accurate to say that before the monster was constructed it was neither a member nor nonmember of the class. The creation of the monster brings with it the question of its kind, and the choice of whether to accept it as legitimate or to exclude it. Both strategies have been adopted by mathematicians. We consider these issues further in Section 11.4.4.

but legitimate counterexamples, Peano came to believe that mathematics must be purged of all use of intuition, which could be accomplished only if all theorems were formally deducible from formally-stated axioms and definitions. To this end Peano also recommended the use of an artificial symbolic logic for mathematical proofs, since natural language could hide unintentional appeals to intuition. The notation developed by Peano, as refined by Russell and Whitehead, is the basis for the notations still used in set theory and symbolic logic. In this Peano acknowledged that he was following in the footsteps of Leibniz, whom he called the "true founder" of logic, and Peano's interest led others to investigate and publicize Leibniz' work on logical calculi.

Peano's Foundations of Mathematics

The reconstruction of the foundations of mathematics by Peano and his colleagues was published in a series of books (1894–1908) and in a journal, which he established for the purpose (1891–1906).[4] Based on Dedekind's work, he reduced analysis to the natural numbers, and the natural numbers to three primitive ideas and five postulates (see p. 222). Thus by the turn of the century he had shown how mathematics could recover its lost rigor. In the process he demonstrated how mathematical propositions could be expressed as symbolic structures, and how intuition could be replaced by deduction implemented through the formal manipulation of these structures.

Interlingua

Like Leibniz, Peano was devoted to the construction of an ideal artificial language to facilitate scientific communication. He deeply regretted the decline of Latin as the international language of science, and so he created *Interlingua*, a dialect of Latin without its complicated inflections, which make it difficult to learn (*latino sine flexione*). The language was, of course, a failure, but Peano worked on it assiduously, and many of his later mathematical publications are written in Interlingua. Thus we see that Peano advocated radical reform in both the *method* and the *medium* of mathematics. In the first he succeeded, but not in the second.[5]

[4] The books were called *Formulaire de Mathématique*, though the last volume was given an Interlingua title, *Formulario Mathematico*. The journal was *Rivista di Matematica* (**1**–**5**), later renamed *Revue de Mathématique* (**6, 7**), and renamed again in Interlingua, *Revista de Mathematica* (**8**).

[5] Peano's linguistic interests were not limited to artificial languages; he also attempted to express in mathematics the grammars of various languages and thereby to find a "universal grammar" underlying them all. In this he anticipated Chomsky (p. 343). However, unlike Chomsky, Peano concluded that there is no universal grammar and that individual

Peano's Axioms

Peano proposed three *primitive ideas* and five *axioms*. The primitive ideas are:

$\mathcal{N}$, intended to be the set of natural numbers,

0, intended to be the number zero, and

$\mathsf{succ}\, n$, intended to be the successor of the number n.

The axioms are (the symbolic logic notation is described on p. 225):

1. 0 is a natural number:
$$0 \in \mathcal{N}.$$
2. The successor of any natural number is a natural number:
$$n \in \mathcal{N} \quad \rightarrow \quad \mathsf{succ}\, n \in \mathcal{N}.$$
3. No two natural numbers have the same successor:
$$m, n \in \mathcal{N} \ \wedge \ \mathsf{succ}\, m = \mathsf{succ}\, n \quad \rightarrow \quad m = n.$$
4. 0 is not the successor of any natural number:
$$m \in \mathcal{N} \quad \rightarrow \quad 0 \neq \mathsf{succ}\, m.$$
5. If a property (a) holds for 0, and (b) holds for the successor of any natural number for which it holds, then (c) it holds for all natural numbers:
$$\begin{aligned} &P(0) \ \wedge \\ &\quad \forall m\{[m \in \mathcal{N} \ \wedge \ P(m)] \rightarrow P(\mathsf{succ}\, m)\} \\ &\qquad \rightarrow \forall n[n \in \mathcal{N} \rightarrow P(n)]. \end{aligned}$$

The last postulate is the all-important principle of mathematical induction. In fact Peano had an additional postulate, but it said only that $\mathcal{N}$ is a class ($\mathcal{N} \in \mathrm{Cls}$).

6.1.2 Logicism

> Mathematics has the completely false reputation of yielding infallible conclusions. Its infallibility is nothing but identity. Two times two is not four, but it is just two times two, and that is what we call four for short. But four is nothing new at all. And thus it goes on and on in its conclusions, except that in the higher formulas the identity fades out of sight.
>
> — Goethe

Introduction

We'll consider briefly the three principal approaches to the philosophy of mathematics, *logicism*, *intuitionism* and *formalism*, and how they led to the theory of computation.[6] Logicism is so called because it attempts to secure the foundations of mathematics by reducing it to logic. The motivation is basically this. Logic is well understood, and few mathematicians disagree about the laws of logic. Further, since the development of formal logic, the laws of logic have been put in a completely precise form. Also, since logic must be used in any mathematical proof, the laws of logic are presupposed by mathematics. Therefore, if we can show that the laws of mathematics reduce to the laws of logic, then we can be as confident in mathematics as we are in logic.

History

Dedekind thought that mathematics is "an immediate product of the pure laws of thought" and sought a "purely logical construction of the science of numbers." Nevertheless, the founders of logicism, Gottlob Frege (1848–1925) and Bertrand Russell (1872–1970), looked to Cantor instead of Dedekind for their definition of number, which we now consider.

Definition of Number

It may not be at all clear that mathematics can be reduced to logic, so we'll give just an indication of how this can be done. If we are to reduce mathematics to logic, then certainly number is one of the concepts that must

grammars are largely determined by chance.

[6] An excellent source for this chapter is Benacerraf & Putnam (*PM*), which contains papers by a number of the originators and promoters of the philosophies of mathematics discussed in this chapter. In particular, pp. 41–65 contain a brief, though technical, statement of the logicist, intuitionist and formalist programs. Less technical accounts are von Mises (MPHU) and the article "Mathematics, Foundations of" in Edwards (*EP*, Vol. 5, pp. 188–213). The comprehensive statement of the logicist program is Russell (*PM*). See also the selections by Frege, Russell and Carnap in Benacerraf & Putnam (*PM*)

be reduced. What is a number? One answer is that a number is a property of collections of things. For example, this pair of apples and that pair of shoes both have the property of being two. Similarly, this triple of sisters and that triple of planets both have the property of being three. Hence, we can see that a number is a property that may belong to *sets* of things: The sets {3, 8} and {Plato, Aristotle} both have the property Two, which we may write:

$$\mathsf{Two}(\{3, 8\})$$
$$\mathsf{Two}(\{\text{Plato, Aristotle}\})$$

Now we must see how the numbers can be expressed in logic. We start with a simpler case, the number 1. Thus we must define a property One that is possessed by all singleton sets (sets with one element). For example, $\mathsf{One}(\{3\})$ and $\mathsf{One}(\{\text{Aristotle}\})$ are both true. How in general can we say that $\mathsf{One}(S)$, that is, that S has exactly one element? First of all we must say that it has at least one element; we can accomplish this by saying that there is an x that is in S. Second we must say that it has at most one element. You can see how to do this by thinking about how you usually prove uniqueness in mathematics. For example, if you know x is a solution to an equation and you want to show it's the unique solution, you assume that there is some other solution y, and then you show that y must equal x. We can do the same here by saying that if y is any element of S then $y = x$. The resulting definition of One can be expressed in symbolic logic as follows:[7]

$$\mathsf{One}(S) \quad \equiv \quad \exists x[x \in S \quad \wedge \quad \forall y(y \in S \ \rightarrow \ y = x)] \tag{6.1}$$

This means that wherever we see the term One we can replace it by that lengthier logical expression.

Just so the idea is clear we consider the property Two. Clearly a set has two elements if there are x and y in S such that x is not equal to y, and any

[7] In this chapter and the next I will make some use of the notation of symbolic logic. Although many of the ideas could be expressed without this notation, there are two important reasons for introducing it here. First, symbolic logic has come a long way since the pioneering efforts of Boole; I don't want you to have the impression that modern symbolic logic is like Boole's system. Second, a major theme of this chapter is the reduction of mathematics to symbolic structures. You will have a much better idea what this entails if you see some examples of the process. A summary of the notation of symbolic logic can be found on p. 225.

Symbolic Logic

These are the principal signs of a typical, modern symbolic logic and their meanings (there are many dialects in use):

$P(x)$	x has the property P
$\neg P$	not P, it is not the case that P
$P \wedge Q$	both P and Q
$P \vee Q$	either P or Q or both
$P \rightarrow Q$	if P then Q
$P \equiv Q$	P if and only if Q
$\exists x\ P$	there exists an x such that P
$\forall x\ P$	for all x, P
$\{x \vert P\}$	the class of all x such that P
$x \in S$	x is in the set or class S

For example, Eq. 6.1 can be read: "S has the property One if and only if there exists an x such that x is in S and for all y in S, y equals x." Similarly, Eq. 6.2 can be read: "S has the property Two if and only if there exist x and y in S such that $x \neq y$ and for all z in S, either $z = x$ or $z = y$."

z in the set is either x or y:

$$\begin{aligned}\mathsf{Two}(S) \;\equiv\; \exists x\, \exists y \;\; \{ \;& x \in S \;\wedge\; y \in S \;\wedge\; x \neq y \\ & \wedge\, \forall z[z \in S \;\rightarrow\; (z = x \vee z = y)]\}. \end{aligned} \tag{6.2}$$

Given the definitions of One and Two we can prove, for example that if we take the union of two disjoint sets with the property One, then we will get a set with the property Two. This is the basis for the truth of $1 + 1 = 2$. However to reach that stage we need to take one further step. Set theory says that for any property there is a set whose members are just the individuals with that property. In other words, the extension of a property is a set (recall p. 119). Therefore, we define the number 2 to be the class of all pairs, that is, the class of all things with the property Two:

$$2 = \{S | \mathsf{Two}(S)\}.$$

Notice that the number 2 is a set of sets. This is to be expected given our view that numbers are properties of collections.

The approach outlined above can be extended to all numbers, which is the essence of the famous definition by Bertrand Russell (1872–1970):

> The number of a class is the class of all those classes that are similar to it.[8]

By *similar* Russell means that the classes may be put in a one-to-one correspondence (a notion which is easy to define in symbolic logic). The idea of a one-to-one correspondence is fundamental to the idea of counting, and hence number (see p. 229).[9]

Russell's Paradox

The method we have illustrated for reducing numbers to logic (the logic of classes in fact) is in simplified form what Gottlob Frege (1848–1925) proposed in his *Fundamental Laws of Arithmetic*, published in 1893–1903. Unfortunately, even before it had been completely published, a fatal flaw had been discovered by Bertrand Russell; it is known as *Russell's Paradox*, and it goes like this. The logic used by Frege stipulates that for any property there is

[8] Russell (*IMP*, p. XX); see also Benacerraf & Putnam (*PM*, p. 172) and Newman (*WM*, p. 542).

[9] Frege's and Russell's discussions of number are reprinted in Benacerraf & Putnam (*PM*, pp. 130–182). See also Russell (*PM*, p. 115) for an earlier statement of the definition of number.

a set of the individuals having that property (i.e., every property has a set that is its extension). Therefore, Russell says, consider the property of not being a member of yourself:

$$Q(S) \equiv S \notin S.$$

For example, since there are more than two pairs, 2 does not belong to the class 2, hence $Q(2)$. Now, consider the class Ω of all things having this property Q; this is the class of all things that are not members of themselves:

$$\Omega \equiv \{S | S \notin S\}$$

The paradox arises from asking whether Ω is a member of Ω. For if it is, then it must satisfy the property Q, which means that it is not a member of itself, which is a contradiction. Conversely, if Ω is not in Ω, then it must satisfy the definition of Q, which means that it is a member of itself. Therefore, the assumption that Ω either is or is not a member of Ω leads to a contradiction.

This is of course a mathematical version of the famous Liar Paradox. Its simplest version is the self-negating statement: "This sentence is false." Paradoxes of this form were known to Aristotle, and the study of them was quite popular in Hellenistic period and the Middle Ages (pp. 61, 64, 76).[10] Russell's contribution was to show that this paradox could be expressed in Frege's logic, and therefore that his logic was inconsistent. Needless to say, this seriously undermined Frege's attempt to secure the foundations of mathematics!

Principia Mathematica

Although Russell had discovered a fatal flaw in Frege's logic, he believed that the logicist program could work. Therefore he designed a logic that did not permit the paradoxical set Ω. He did this by defining a type theory which permitted propositions of the form '$x \in S$' only when the type of x was "lower" than that of S. This rule prohibited propositions of the form $S \in S$; in effect it said that it makes no sense to even ask whether a set is a member of itself. The trouble with Russell's type theory is that it's quite complicated, and nowhere near as intuitive as the usual laws of logic (such as 'A and B implies A'). Nevertheless, Russell, in collaboration with Whitehead, set out to reduce all of basic mathematics to logic. The result of this project is the massive *Principia Mathematica*, three volumes of dense symbolic logic. The necessity of the type theory considerably complicates the proofs, and we must get to page 83 of volume 2 before $1 + 1 = 2$ is proved![11]

[10] A good collection of relevant texts is Bocheński (*HFL*, §23).

[11] Specifically, it is Theorem *110.643; see Russell & Whitehead (*PM*).

Furthermore, Russell discovered that the usual laws of logic were not sufficient to prove the expected mathematical theorems. For example, to prove that the infinite set of integers exists, he had to postulate an additional Axiom of Infinity. This again was much less intuitive than the familiar axioms of logic. Furthermore, to show the existence of the real numbers he needed an additional axiom.

Failure of Logicism

Needless to say, mathematicians saw little advantage in trying to secure the foundations of mathematics by reducing arithmetic, which is very intuitive, to Russell's logic, which is anything but. Therefore, the logicist program was a technical success (at least if you believe Russell's axioms), but it was a political failure; it failed to win the allegiance of mathematicians.

6.1.3 Intuitionism

> God made the integers; all else is the work of man.
>
> — Leopold Kronecker

Cantor's Set Theory

Georg Cantor (1845–1918) did much to bring mathematics to its modern form, but his accomplishments have not been universally appreciated. Indeed, few mathematicians have been so controversial as Cantor, and many mathematicians find his set theory to be outrageous. The reason is this. For millenia mathematics had been trying to rid itself of the infinite, and, as we've seen, it had largely succeeded. But Cantor claimed that the infinite was a proper subject of mathematical study, and proceeded to invent a theory of "transfinite numbers" (i.e. numbers beyond the finite). So that you can understand how such a thing is possible, we'll sketch the barest outlines of Cantor's theory.[12]

Transfinite Numbers

Cantor begins with one of the most basic concepts of arithmetic, the idea of a *one-to-one correspondence*. This is the principal that underlies counting. To see if two collections of things have the same number of elements, we line them up; technically, we set up a one-to-one correspondence (Fig. 6.1). If two sets can be made to correspond one to one, we say that they have the same *number* of elements. Indeed, we can define a number to be a class of sets that all have the same number of elements in this sense (i.e., the definition Russell

[12]More information on transfinite numbers can be found in the article "Set Theory" in Edwards (*EP*, Vol. 7, pp. 420–427). See also Hausdorff (*ST*), especially Chapter II.

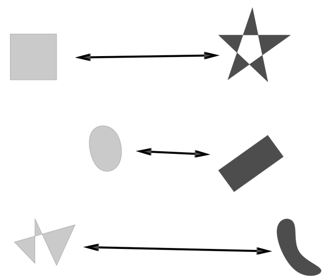


Figure 6.1: One-to-one Correspondence. If two sets can be made to correspond in a one-to-one fashion, then we say that they have the same number of elements.

(later) gave, p. 223). This is all routine, so long as the sets involved are finite. Cantor's contribution comes in extending these ideas to infinite sets. Thus he says that all sets that can be put into a one-to-one correspondence with the integers have the same number of elements, and he calls this infinite number $\aleph_0$ (aleph nought). Similarly, all sets that can be put in a one-to-one correspondence with the reals, have the same number, which Cantor writes $\mathbf{c}$, the number of the continuum. Further, Cantor showed that these two transfinite numbers are different, $\aleph_0 \neq \mathbf{c}$, and in fact $\aleph_0 < \mathbf{c}$ (see p. 233).

Now, it's reasonable to ask how many of these infinite numbers there may be, and Cantor answers this question. In particular, Cantor shows that a set cannot be put into a one-to-one correspondence with all its subsets, so the set of all these subsets has a larger number of elements than the set. In particular, if a set has X elements, then the set of its subsets has 2^X elements. The remarkable thing is, that this holds even if the set has an infinite number of elements. Thus the number of subsets of the integers is $2^{\aleph_0}$, an infinite number which Cantor calls $\aleph_1$. It is easy to see that there is an endless series of infinite numbers, each larger than the previous: $\aleph_0, \aleph_1, \aleph_2, \ldots$ where $\aleph_{k+1} = 2^{\aleph_k}$. This is surely a remarkable idea.

The Continuum Problem

A natural question, once one has accepted the transfinite numbers, is whether $\aleph_0, \aleph_1, \ldots$ are all the transfinite numbers. In particular, we may ask whether $\mathbf{c}$ occurs in this list. Cantor answers this question in the affirmative by proving $\mathbf{c} = \aleph_1$. Another obvious question is whether there is an transfinite number between $\aleph_0$, the number of integers, and $\aleph_1$, the number of reals. Cantor was unable to answer this question, but conjectured that there was no such number, and determining the truth of this *continuum hypothesis* was the first of the 23 unsolved mathematical problems listed by Hilbert in 1900 as the most pressing issues for twentieth-century mathematics. The problem has now been solved, but in a surprising way. In 1938 Kurt Gödel (1906–1978) showed that the continuum hypothesis is consistent with the most widely accepted axioms of set theory,[13] and therefore that it cannot be *disproved* from those axioms (assuming that they are consistent). In 1963 Paul J. Cohen proved the complementary result, that the denial of the continuum hypothesis is also consistent with these axioms, and therefore the hypothesis cannot be *proved* from them (again assuming their consistency). Since the continuum hypothesis can be neither proved nor disproved from these axioms, it is *independent* of them, and we can assume it or not as we like. Incidentally, this also shows that the standard axioms of set theory are *incomplete*, a topic about which we will have much more to say later (Sect. 6.1.4, Ch. 7).[14]

Criticism of Cantor's Theory

Although Cantor's theory of transfinite numbers has been accepted into the mainstream of mathematics, it has not been without its critics, both now and in the past. Many important mathematicians, such as Kronecker, Borel and Weyl, believed that Cantor's theory was no more meaningful than chicken scratches on the ground.[15] Just because the symbols could be manipulated without encountering a contradiction, it did not follow that the symbols *meant* anything. Mathematics appeared to be in danger of breaking loose from its moorings in common sense and drifting off into a Never-Never Land of fantastic inventions. But was there some *principle* of mathematics by which Cantor's set theory could be rejected? Although several mathematicians formulated such principles, Brouwer's *intuitionism* was the most influential.[16]

The Intuitionist Viewpoint

Intuitionism in effect says this: There is nothing special about logic and it has no priority over mathematics. In fact the laws of logic are abstractions of

[13] That is, the Zermelo-Fraenkel axioms without the axiom of choice.

[14] The independence of the continuum hypothesis is not a peculiarity of the Zermelo-Fraenkel axiomatization; it holds for a variety of axiomatizations of set theory. For more on the continuum problem, see Gödel (WCCP) and Cohen (*STCH*); a useful discussion of the result's significance is Smullyan (CP).

[15] Leopold Kronecker: 1823–1891; Émil Borel: 1871–1956; Hermann Weyl: 1885–1955; L. E. J. Brouwer: 1882–1966.

[16] Relevant papers by Brouwer and Heyting are reprinted in Benacerraf & Putnam (*PM*). See also Kneebone (*MLFM*, Ch. 9) for a good overview of intuitionism.

what mathematicians do when they prove things. Therefore logic is based on mathematical practice. Rather than trying to reduce mathematics to logic, we should instead go back to the primary intuitions upon which mathematics is based. When we do this, we find that mathematics is based on the simplest of possible ideas, for the number series 1, 2, 3, ... is nothing more than the idea of succession. Intuitionism takes this basic idea and the simple arithmetic reasoning that accompanies it as the basis for mathematics. These are the primaries to which valid mathematics must be reducible.

Constructive Proofs

Mathematicians have traditionally preferred constructive proofs; it is generally more useful to actually have a number (or a way of computing it), than to just have a proof that it exists. Nevertheless, nonconstructive proofs have been accepted when constructive proofs were either unavailable or much more complicated than the nonconstructive. The degree of emphasis on constructive proofs had largely been a matter of mathematical style, but then Cantor's set theory appeared on the scene. To mathematicians with constructivist convictions it was anathema. They argued that Cantor's infinite numbers could not be constructed in any reasonable way. Mathematicians such as Kronecker and Weyl thought that this was the wrong direction for mathematics, so they attempted to reformulate it along completely constructive lines. This program was pursued most aggressively by Brouwer, the founder of intuitionism.

Primary Intuition

For intuitionists all our mathematics stems from a primary understanding of the natural number series, specifically, from our ability to mentally construct each element of the series from the preceding. Therefore, all mathematical objects must be constructed by these same mental operations. We cannot claim to know that a mathematical object exists unless we know how to construct it. Traditionally mathematicians had often used indirect existence proofs: showing that the nonexistence of a thing would be contradictory. Such proofs were unacceptable to the intuitionists, for they did not permit an intuitive apprehension (construction) of the object claimed to exist. Therefore, for intuitionists, constructive proofs are the only valid existence proofs.

Positive Existence Proofs

We have seen that an intuitionist (or constructivist) proves the existence of a number (or other mathematical object) by showing how to construct it. Often, however, mathematicians want more than a single number; they want an answer for each problem in some class. For example, they might want to show that for all equations of a certain type, there is a number that's the solution to the equation. In these cases the intuitionist must give a general procedure that can be shown to allow the construction of the answer to each such problem.

Negative Existence Proofs

A special problem arises with negative existence claims, such as "there is no number such that ..." In general it would seem that to prove such a proposition we would have to run through all the numbers, showing that each does not have the property in question. The intuitionist avoids this infinite process by showing that the existence of a number having the property would lead to a contradiction. This is done by exhibiting a construction process that allows the derivation of a contradiction, such as $0 = 1$, from any number having the property. Primary intuition of the number series tells us that $0 = 1$ false, so the existence of the number can be rejected.

Rejection of Law of Excluded Middle

The intuitionist idea of mathematical proof leads to some interesting logical consequences. For example, consider Fermat's Last Theorem, which states that $X^n + Y^n = Z^n$ has no integer solutions for $n > 2$. An intuitionist proof of this theorem would have to show that the existence of such a solution would lead to a contradiction; this has never been done. On the other hand, an intuitionist *disproof* of this theorem would have to actually construct the X, Y, Z, and n that solve the equation; this also has not been done. Therefore, from an intuitionistic viewpoint the theorem is *neither true nor false*, since it has been neither constructively proved nor disproved. More carefully, it makes no sense to talk of the existence of a solution unless it's been constructed; it makes no sense to talk of the nonexistence of a solution unless the appropriate contradiction has been derived. Therefore, if neither has been accomplished, then it makes no sense to talk of the theorem being true or false. The intuitionist view rejects a logical principle that goes back to Aristotle: The Law of the Excluded Middle. This principle says that either a proposition or its negation must be true; there is no third alternative. Formally, $A \vee \neg A$.[17] Therefore, if we've shown A cannot hold, then we can conclude not-A does hold. From the intuitionist perspective, a fact that A does not hold tells us nothing about the truth of not-A. For it could be the case both that there is no solution to the equation, and that we have insufficient means to derive a contradiction from a supposed solution. In this case neither A nor not-A is true.

Limitations on Rejection

It is widely believed that intuitionists completely reject the Law of the Excluded Middle. This is not so. The use of the law agrees with primary intuition, so long as we're dealing with finite sets. It's only when we're

[17]That is 'A or not-A'; see p. 225 for the notation of symbolic logic.

dealing with infinite sets, and it becomes impossible, even in principle, to inspect all the elements, that it is problematic. In this case we have the third possibility: we can neither find a solution nor show that its existence is contradictory.

Influence of Intuitionism

The intuitionist program was at one time appealing to many mathematicians, but its detailed elaboration proved too complicated and ultimately — unintuitive. For example, Weyl (*PM&NS*, p. 54) said,

> Mathematics with Brouwer gains its highest intuitive clarity. He succeeds in developing the beginnings of analysis in a natural manner, all the time preserving the contact with intuition much more closely than had been done before. It cannot be denied, however, that in advancing to higher and more general theories the inapplicability of the simple laws of classical logic eventually results in an almost unbearable awkwardness. And the mathematician watches with pain the larger part of his towering edifice which he believes to be built of concrete blocks dissolve into mist before his eyes.

This was in part due to the rejection of classical logic, but also in part a result of the insistence on constructive proofs. Thus, although intuitionism was a tonic, and made mathematicians more wary of nonconstructive methods, it failed to win their allegiance. Nevertheless, as we'll see, some intuitionist ideas were adopted into the formalist movement, and ultimately led to the theory of computation (see p. 235).[18]

Nondenumerability of the Reals

Because it will illuminate the important proofs of Gödel and Turing discussed in Ch. 7, we sketch here Cantor's *diagonal proof* that the real numbers are *nondenumerable*, that is, that they cannot be put in a one-to-one correspondence with the natural numbers. First note that infinite sets have many unintuitive properties. For example, one set may be a part of another, yet have the same (transfinite) number of elements. For example, the set of even natural numbers is a subset

[18]Interestingly, intuitionist ideas are still finding application in computer science, which should not be surprising, since computer science is concerned with the computable, and intuitionism has attempted to reformulate mathematics constructively. For example, intuitionist set theory has been applied to the theory of types in programming languages.

of the set of all natural numbers, yet they have the same number $\aleph_0$ of elements because they can be put in a one-to-one relation, $2n \longleftrightarrow n$. Even the number of rationals is $\aleph_0$, as we can see by enumerating them in the special order:

$$\frac{1}{1},\quad \frac{2}{1},\quad \frac{1}{2},\quad \frac{1}{3},\quad \frac{2}{2},\quad \frac{3}{1},\quad \frac{1}{4},\quad \frac{2}{3},\quad \frac{3}{2},\quad \frac{4}{1},\quad \ldots$$

(First those whose numerator and denominator sum to 2, then those which sum to 3, etc.) Every rational number will eventually appear in this list.

The question that naturally arises is whether the number $\mathbf{c}$ of reals is also $\aleph_0$. Cantor uses a proof by contradiction to show it's not. For suppose we had some enumeration of all the real numbers between 0 and 1, expressed here as (possibly infinite) decimal expansions (arbitrary examples):

$$\begin{array}{cccccccc}
1 & \longleftrightarrow & 0. & \boxed{3} & 7 & 7 & 2 & \ldots \\
2 & \longleftrightarrow & 0. & 0 & \boxed{9} & 9 & 4 & \ldots \\
3 & \longleftrightarrow & 0. & 5 & 0 & \boxed{6} & 7 & \ldots \\
4 & \longleftrightarrow & 0. & 2 & 5 & 1 & \boxed{2} & \ldots \\
\vdots & & \vdots & & & & &
\end{array}$$

The digits on the diagonal of the table have been marked. Now write down a (possibly infinite) decimal expansion that differs from the number on the diagonal. That is, for the first digit pick something other than 3, for the second, something other than 9, and so forth; we might get:

$$Q \;=\; 0.\;\; 4 \;\; 8 \;\; 3 \;\; 3 \;\; \ldots$$

We claim that this number Q cannot appear anywhere in our (supposedly) exhaustive enumeration of the reals in $(0, 1)$, since Q differs from the nth number in its nth digit. Since this conclusion contradicts our assumption that we can have an enumeration of all the reals between 0 and 1, we must reject that assumption and conclude that we cannot. (There is a small hole in this proof that is easy to plug: one real may have two different decimal expansions, e.g., 0.2000… and 0.1999….)

6.1.4 Formalism

bert: 1862–1943

Formalism can be seen as an attempt to have the best of logicism and intuitionism.[19] On the one hand it rejects intuitionism's extremely limited proof techniques. As David Hilbert said in 1925, "No one shall drive us out us of the paradise which Cantor created for us."[20] The formalists wanted to permit mathematicians to use any axioms and rules of inferences so long as they did not lead to a contradiction. But Russell's paradox had shown how easily inconsistency can creep into an axiomatic system. Therefore the formalists were concerned with *proving* the consistency of their axiomatic systems, but to be a subject of mathematical proof, an axiomatic system must be treated as a mathematical object. This allows us to prove theorems about the proofs and theorems of this system. That is, we are doing *metamathematical* proofs about mathematical systems (axiomatic systems).

Completeness and Independence

Though consistency is the *sine qua non* of formal axiomatic systems, additional characteristics such as *completeness* and *independence* are also desirable. An axiomatic system is *complete* if it's possible to either prove or disprove every proposition that can be expressed in the language of the axiomatic system. In effect completeness means that we have enough axioms to completely characterize the subject matter of the axiomatic system. An axiom is *independent* of the others if it cannot be derived from them; therefore independence addresses the absence of redundant axioms. Metamathematical methods are also used to establish the completeness and independence of formal axiomatic systems.

Consequences of Formalism

The formalist approach has two important consequences. First it means that axiomatic systems must be truly formal; there can be no intuitive appeals in proofs if they are to be treated as mathematical objects. Second, it means that if the resulting proofs of consistency, completeness and independence are to be convincing, then metamathematical proofs must be simpler and more intuitive than those permitted in the axiomatic systems that are their objects. Here the formalists followed the intuitionists, for they stipulated that metamathematical proofs should be purely finitary and constructive. In effect they agreed with the intuitionists that these techniques are

[19]Original papers by Hilbert and von Neumann are reprinted in Benacerraf & Putnam (*PM*). The definitive exposition of formalism is Hilbert and Bernay's *Grundlagen der Mathematik* (*Foundations of Mathematics*) (2 vols., 1934, 1939).

[20]See Benacerraf & Putnam (*PM*, p. 191); cf. Hilbert, *Math. Ann.*, 95, 1926, p. 170, and *Grundlagen der Geometrie* (*Foundations of Geometry*), 7th ed., 1930, p. 274.

as secure as anything we know, so that they should be used for establishing the consistency, completeness and independence of more powerful axiomatic systems. In this section we will discuss the resulting idea of a formal system; in Chapter 7 we will discuss the outcome of the metamathematical investigations. First however we'll consider the formalist approach in more detail.

6.2 Interpretations and Models

Substantial Axiomatic Systems

To understand the formalist program it's necessary to realize that it introduces a new idea of an axiomatic system. The traditional idea of an axiomatic system, deriving from Aristotle (Section 2.5.3, p. 49), took it to be a reduction of a body of empirical knowledge to a set of self-evident axioms and definitions, that is, to primary truths that provide the evidence for themselves. Hilbert and Bernays called such an axiomatic system *inhaltlich* (substantive or contentful) since its axioms and propositions are about some body of experience or some things in the world (e.g., physical space or time, objects moving under the influence of forces). The intuitionist and logicist accounts are of this kind: intuitionism takes the propositions of mathematics to be about primary intuitions of sequence; logicism takes them to be about properties of classes and other logical objects.

Formal Axiomatic Systems

In contrast to the traditional view, the formalists believed that although empirical content might affect the practical relevance of a mathematical theory, it had no bearing on its mathematical truth, which depends only on its logical soundness. A *formal* axiomatic system can be defined by setting down any consistent set of axioms. The resulting system of theorems express truths about a class of abstract structures, which might or might not refer to anything in the world or experience. For example, Euclidean geometry had been traditionally understood as a substantive axiomatic system expressing truths about the nature of space. When non-Euclidean geometries were discovered in the nineteenth century, mathematicians realized that it is empirical question whether physical space is in fact Euclidean or non-Euclidean. Interpreted as a formal axiomatic system, Euclidean geometry defines a class of abstract structures (Euclidean spaces), which might or might not include the physical space in which we live. In effect the formalists divided the mathematical investigation of formal Euclidian geometry from the substantive, scientific investigation of the applicability of these axioms to physical space.[21]

Interpretations

Formal axiomatic systems are normally taken to be *about* something. That is, although for the purposes of logic the symbols are treated as meaningless tokens, our ultimate goal is to interpret the symbols as denoting objects in some domain. For example, if we interpret Peano's axioms (p. 222) in the obvious way, then we will interpret the symbol '0' as meaning the number zero, interpret formulas of the form '$\mathsf{succ}(n)$' as meaning the successor of n, and interpret propositions of the form '$X = Y$' as meaning that X and Y are the same number. This is the *intended interpretation* of Peano's axioms.

An axiomatic system is *interpreted* by setting up a correspondence between symbols in the system and objects in the *domain of interpretation*.[22] For an example, take the domain to be the natural numbers. Then corresponding to each constant symbol of the language (such as '0') there is an individual object of the domain (such as the number 0). Corresponding to each function symbol (such as 'succ') there is a function on the objects of the domain (such as the successor function). And corresponding to each predicate or relation symbol (such as '=') there is a predicate or relation on the domain (such as numerical equality). Thus every proposition expressible in the language of the axiomatic system corresponds to some statement about the domain of interpretation. These statements may be true or false.

Models

Clearly, there is no guarantee that the theorems of an axiomatic system will make true assertions about a domain. However, if an interpretation is such that the axioms are true in the domain, and the inference rules of the system are logically valid, then it will follow that the theorems are also true about the domain. This is of course the intended situation, and in this case the interpretation is called a *model* or *realization* of the axiomatic system.

Syntax vs. Semantics

The interpretation of formal systems illustrates the distinction between

[21] The foregoing should not be interpreted as a claim that mathematicians are or should be uninterested in the scientific application of their theories — though some mathematicians have taken that view. The well-known and highly-respected mathematician G. H. Hardy (1877–1947; see Hardy, MA, p. 2038) said, "I have never done anything 'useful.' No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world." On the other hand, many of the best mathematicians have stressed the importance of applications for the health of their discipline. The formalist position is addressed to the ultimate standard of mathematical truth, not to practical questions of the particular truths to be investigated.

[22] The theory of interpretations of axiomatic systems was developed in the 1930s by Alfred Tarski (1902–1983); see Tarski (*IL*, Ch. VI) for a readable introduction, the original paper is Tarski (CTFL).

syntax and *semantics*. By design, the deductive or derivational processes of a formal axiomatic system are *purely syntactic*, that is, they depend only on the shapes and arrangements of the symbols, not on their meaning, an idea which goes back to Aristotle's formal logic (Section 2.5.4). Nevertheless, it is often useful to associate meanings with the symbols and formulas of a formal system, and such an association is a *semantics* for the system. If the association is done correctly, as in a model, then the syntactically permitted derivations will mirror semantically sound inferences. This is the principal advantage of formal systems, which makes them so attractive in mathematics, but also in cognitive science and artificial intelligence. For by mirroring semantic processes in syntactic manipulations, formal systems show the way to accounting for inference in purely physical terms.[23]

Consistency

The formalist philosophy of mathematics tries to allow the mathematician the maximum flexibility. Axiomatic systems are not required to be simply logic (as in logicism), or to be justifiable by appeals to constructions from primary intuitions (as in intuitionism). To the formalist, axiomatic systems are legitimate so long as they are internally consistent. In syntactic terms, consistency simply means that there is no proposition P expressible in the system such that both P and not-P are provable from the axioms by the rules of inference. It is easy to show that every proposition is provable in an inconsistent formal system supporting the usual laws of logic. That is, each expressible proposition and its negation are both provable in the system. Clearly, there is not much point in having proofs at all if everything is provable. This is why consistency is the *sine qua non* of formal systems.

Decidability and Completeness

We've seen that for any proposition P it's important that at most one of P and not-P be provable. This suggests that it may be possible that *neither* P nor not-P is provable. This would be the case if the axiomatic system didn't have sufficient axioms to either prove or disprove every proposition expressible in the system. Such a system is called *incomplete*. We can state the matter more precisely as follows. Call a proposition P *decidable* in the system $\mathcal{A}$ if either P or not-P is provable from the axioms by the rules of inference. Then the formal axiomatic system $\mathcal{A}$ is *complete* if every expressible proposition in the system is decidable in that system. Thus, for

[23] The existence of purely syntactic yet interpretable systems is also the basis for practical computing; it allows us to manipulate physical quantities such as electric charges, yet interpet them as quantities of money, people, etc.

Everything Provable From Inconsistency

An axiom tells us that $P \rightarrow P \vee Q$ for any Q, hence by the definition of '$\rightarrow$' we conclude $\neg P \vee P \vee Q$. By DeMorgan's law this is $\neg(P \wedge \neg P) \vee Q$, which may be rewritten as an condition:

$$(P \wedge \neg P) \rightarrow Q$$

Thus a contradiction $P \wedge \neg P$ implies any other arbitrary proposition Q. (The notation of symbolic logic is summarized on p. 225.)

the formalists, axiomatic systems should be both complete and consistent, in other words, *exactly one* one of P and not-P should be provable.[24]

[24]Compared to consistency and completeness, independence may seem something of a luxury, but it's closely related to consistency. For example, if we can prove that any proposition expressible in the system is independent of the axioms, then we know that the axioms are consistent, since everything is provable from inconsistent axioms.

Semantic Consistency and Completeness

Semantic Completeness

We have defined consistency and completeness in purely syntactic terms: a formal axiomatic system is consistent and complete if, considered as a mechanical formula generating process, it can generate, for any formula P, either P or not-P, but not both. There are also closely related semantic notions of consistency and completeness. We can say that an interpreted formal system is *semantically complete with respect to an interpretation* $\mathcal{I}$ if it permits every *true* expressible proposition to be proved. It turns out that semantic completeness with respect to $\mathcal{I}$ is equivalent to syntactic completeness if $\mathcal{I}$ is a model.

To see this, first suppose the system is semantically complete with respect to $\mathcal{I}$. If P is any expressible proposition, then in $\mathcal{I}$ either P is true or not-P is true. Whichever is true must be provable, since the system is semantically complete with respect to $\mathcal{I}$. Since as a result either P or not-P is provable, we conclude that the system is also syntactically complete.

Conversely, suppose the system is syntactically complete, but semantically incomplete with respect to $\mathcal{I}$. Since it's semantically incomplete there is some true proposition P that cannot be proved. On the other hand, since not-P is false in model $\mathcal{I}$, it's also unprovable. Thus in this case neither P nor not-P is provable, which contradicts the assumption that the system is syntactically complete.

Semantic Consistency

There is also a notion of *semantic consistency* with respect to an interpretation, which means that every provable proposition is true in $\mathcal{I}$ (notice that this is the converse of semantic completeness). It also turns out that semantic consistency with respect to $\mathcal{I}$ is equivalent to syntactic consistency if $\mathcal{I}$ is a model.

To see this, first suppose that the system is semantically consistent with respect to $\mathcal{I}$. Obviously it must also be syntactically consistent, since if it weren't, then every proposition would be provable, and hence true in the model $\mathcal{I}$, which is impossible.

Conversely, suppose that the system is syntactically consistent. By the definition of models, any conclusions drawn must also be true of the model $\mathcal{I}$. That is, any provable proposition is true in $\mathcal{I}$, and so the system is semantically consistent with respect to $\mathcal{I}$.

> Since we are most interested in logically valid axiomatic systems and we can often verify the truth of the axioms for an intended interpretation, for these interpretations semantic consistency and completeness are equivalent to syntactic consistency and completeness, again demonstrating the important mirroring of semantics by syntax. With respect to a model for a semantically consistent and complete axiomatic system, a proposition is provable if and only if it's true — clearly a desirable situation.

6.3 Metamathematics

Metamathematics of Geometry

Hilbert first demonstrated his metamathematical methods in 1899 in his *Grundlagen der Geometrie* (*Foundations of Geometry*), which proved consistency and independence results for his axioms for Euclidean geometry. The method he used was to invent models for the axiomatic system. In general, he interpreted geometric propositions as propositions about real numbers by the use of Cartesian coordinates. First he showed the consistency of his axioms by interpreting the symbols in the usual, intended way (e.g. lines are interpreted as solutions of linear equations). Thus an inconsistency in the geometry would imply an inconsistency in arithmetic. Further, if the axioms for geometry were inconsistent, then everything would be provable from them, and so likewise we would have to conclude that everything is true in arithmetic, including such obviously false propositions as $0 \neq 0$. In a similar way Hilbert showed the independence of an axiom A_1 from the other axioms $A_2, \ldots, A_n$ by constructing a nonstandard interpretation in which A_1 is false but $A_2, \ldots, A_n$ are true. For if A_1 were not independent, it would be provable from $A_2, \ldots, A_n$ and so true, thus contradicting its falsity.

Relative Consistency

It should be apparent that Hilbert's method can establish the absolute consistency of geometry only if we know there are no contradictions in arithmetic, for if there are contradictions, then everything is provable from them, including $0 \neq 0$. Similarly, the independence proofs presuppose the consistency of the nonstandard interpretations, since if they are inconsistent then they will permit A_1 to be false and simultaneously derivable from $A_2, \ldots, A_n$ and so true. In fact, Hilbert's *Grundlagen der Geometrie* did not establish the *absolute* consistency of his axioms, but only their consistency *relative* to the consistency of (real number) arithmetic, which was hardly unproblematic. Thus Hilbert was obliged to look further for a noncontradictory bedrock on which to build mathematics.

Need for a Secure Arithmetic

Though Hilbert's metamathematics of geometry is a remarkable accomplishment, it is also ironic, for by using arithmetic as a model for geometry it exactly reverses Descartes' use of geometry to justify arithmetic, a reversal which shows the the strength of the arithmetizing movement. Nevertheless, the development of analysis, which had accomplished the arithmetizing of geometry, had made use of sophisticated mathematical devices, such as infinite sequences and Cantor's set theory, which could hide inconsistencies. The full axiomatic arithmetic of Peano could not be guaranteed to be a secure foundation for mathematics.

Finitary Arithmetic

Certainly there is a core of basic arithmetic — natural number calculation — which is as secure as we could hope. This is the arithmetic that Pythagoras would recognize, wherein numbers are represented by finite arrangements of definite tokens and arithmetic operations are accomplished by physically possible mechanical manipulations. In effect, the intuitionist chose to limit themselves to this mathematics, though they permitted themselves the freedom of adopting other, equally intuitive procedures when the need arose. Unfortunately, the intuitionists themselves showed the difficulties attending this route, and Hilbert along with the majority of mathematicians preferred to remain in their Paradise. The formalist alternative was ingenious: instead of using secure, finitary inference to do mathematics, as did the intuitionists, they would use finitary inference to reason about mathematical systems that permitted more convenient nonfinitary inference. However, accomplishing this requires mathematical reasoning to be reduced to calculation — the manipulation of calculi — of the same transparency as Pythagorean arithmetic.

Metamathematics of Arithmetic

We will not consider here the means by which the formalists reduced mathematical inference to calculation, though a small formal proof is shown on p. 243, and the general method will become clear enough in the remainder of this chapter and the next. Here I will merely summarize what Hilbert's metamathematics was able to establish.

Consistency of the Restricted Predicate Calculus

The main accomplishments of the formalist program are set forth in the two volumes of Hilbert and Bernay's *Grundlagen der Mathematik* (*Foundations of Mathematics*, 1934, 1939). They begin by showing the consistency

Example of Formal Proof

We exhibit here a short proof in logic (adapted from Mendelson, *IML*, pp. 31–32), so that the reader unfamiliar with formal mathematics can get some idea of the nature of a purely formal proof. The main thing to be observed is that the process is *purely syntactic*, that is, it is accomplished by manipulating tokens according to their shape and spatial arrangement. *It is not necessary to understand the symbols in order to check the validity of the proof.* The logical system, the *propositional calculus*, has the following three axiom schemes:

A1. $(P \to (Q \to P))$
A2. $((P \to (Q \to P)) \to ((P \to Q) \to (P \to Q)))$
A3. $(((\neg Q) \to (\neg P)) \to (((\neg Q) \to P) \to Q))$

These are "axiom schemes" because an axiom results from substituting any propositions for the variables (P, Q, R). In addition to the axiom schemes the system has just one *rule of inference* or *derivation*, called *modus ponens* (MP):

$$\text{MP.} \quad P, (P \to Q) \ \vdash \ Q$$

Whenever we have derived formulas of the forms given on the left-hand side of the '$\vdash$' (derives) symbol, we are permitted to derive the corresponding formula given on the right-hand side. The theorem to be proved is: '$(A \to A)$'. Here is a formal proof; notice that each step is either the use of an axiom (indicated by A1, A2, A3) or the derivation of a formula by the application of modus ponens (indicated by MP(m, n) where m, n are the numbers of the steps to which it's applied).

1	$((A \to ((A \to A) \to A)) \to ((A \to (A \to A)) \to (A \to A)))$	A2
2	$(A \to ((A \to A) \to A))$	A1
3	$((A \to (A \to A)) \to (A \to A))$	MP(1, 2)
4	$(A \to (A \to A))$	A1
5	$(A \to A)$	MP(3, 4)

of their formal logic (the restricted, or first-order, predicate calculus). They did this by proving that the axioms possess a certain property (intuitively corresponding to their truth) and that the rules of inference preserve this property. However, since one can easily show that some propositions (the intuitively false ones) do not have this property, it follows that these propositions are not provable, and so the system is consistent (since everything is provable in an inconsistent system).

Consistency of Peano Axioms

Having established the consistency of their logical instrument, they proceeded to prove the consistency of a series of progressively more powerful axiomatizations of natural number arithmetic. The basic method is the same: prove that there is some property that every provable proposition inherits from the axioms, but that is not possessed by some proposition, such as $0 \neq 0$. The unprovability of even one proposition shows consistency. By this method they were able to show the consistency of a system that included all of Peano's axioms (p. 222). They also showed that this system was complete in the sense that all propositions about numbers were decidable (either provable or disprovable) in the system.[25]

Consequences of Recursive Definition

One might suppose that with the establishment of the consistency and completeness of Peano's axioms the formalists had found their secure foundation. However, these axioms are not a sufficient basis for mathematics, since Peano had also used recursive definition. For example, he defined natural number addition by these equations:

$$\begin{aligned} m + 0 &= m, \\ m + \mathsf{succ}\, n &= \mathsf{succ}(m + n), \end{aligned}$$

where $\mathsf{succ}\, n$ represents the successor of a number ($\mathsf{succ}\, 0 = 1, \mathsf{succ}\, 1 = 2$, etc.). Similarly, multiplication is defined:

$$\begin{aligned} m \times 0 &= 0, \\ m \times \mathsf{succ}\, n &= (m \times n) + m. \end{aligned}$$

Hilbert and Bernays were able to show that when the recursive definition of addition was added to the Peano axioms the result was still consistent and complete. Unfortunately this system is still quite weak, since it is incapable of expressing such fundamental mathematical ideas as prime numbers. In contrast, adding multiplication makes much interesting mathematics accessible

[25]Specifically, they established completeness for propositions not containing propositional or predicate variables.

to the axiomatic system, but Hilbert and Bernays were *not* able to establish the consistency or completeness of this system. The reasons go very deep into the nature of symbolic knowledge representation, and will occupy us in Chapter 7. In the meantime we will consider how the formalists' techniques led to the development of the theory of computation — decades before the first computers.

6.4 Models of Computation

6.4.1 Introduction

Formalization of Effective Calculability

The reduction of natural-number arithmetic to calculation — the manipulation of calculi — had been understood since the time of Pythagoras. It was also well known how this method could be extended to real numbers for which a calculable convergent sequence was known (recall Theon's algorithm for the square root of 2, Fig. 2.7, p. 32). In all these cases the calculation has been reduced to an *algorithm*, a precisely specificed procedure that can be completed in a finite amount of time.[26] Thus by the early twentieth century the effective calculation of numbers was well understood. However, the formalist program required *reasoning* to be reduced to effective procedures, that is, it required inference to be reduced to calculation. The ground work had been laid by Aristotle, and algorithms for deduction were demonstrated by Leibnitz, Jevons and others, but metamathematics required that these algorithms be specified so precisely that they could become the objects of mathematical investigation. Therefore, pursuit of the formalist program led to attempts to define precisely such intuitive concepts as algorithm, formal system, computation, and effective calculability.

History

A number of alternative definitions of effective calculability were proposed in the early twentieth century, especially between 1920 and 1940. These include Post production systems (1920s), the lambda calculus (1932), combinatory logic (mid 20s–mid 30s), recursive functions (1936), Turing machines (1936) and Markov algorithms (1954).[27] They have in common that they

[26]Although it's not possible to calculate the entire decimal expansion of the square root of two in finite time, it is possible to calculate it to any specified number of decimal places in finite time, for example, by Theon's algorithm. In this sense the entire decimal expansion is *potentially* computable.

[27]Many of the basic papers describing these models are reprinted in Davis (*Undec.*).

were trying to model formula manipulation procedures that can be accomplished with pencil and paper using a finite amount of effort, for all rigorous mathematical reasoning would seem to fit this description. Nevertheless, the models were quite different from one another, so it is all the more remarkable that they have been proved equivalent. That is, anything that can be computed with Turing machines can also be computed with Markov algorithms and vice versa, and anything computable by Markov algorithm is computable in the lambda calculus, and so forth. Thus all these models lead to the same notion of computability, which is most often called *Turing computability*, after Alan Turing (1912–1954), the inventor of the Turing machine model. The idea of Turing computability is now so well understood that new, equivalent models of computations can be produced upon demand. Indeed, every new computer architecture or programming languages amounts to a new model of Turing computability.

Church's Thesis

The question often arises of whether Turing computability is equivalent to the intuitively computable or effectively calculable. Although Turing computability is precisely defined (by any of the models listed above), effective calculability is not, and so formal proofs cannot be used to establish the identity of the two ideas. However, attempts to formalize the notion of effective calculability inevitably lead to models of computation that are equivalent to Turing computability, which provides strong circumstantial evidence that the ideas correspond. Alonzo Church (1903–), the inventor of the lambda calculus, was the first to suggest that effective calculability be identified with

---

Emil Post (1897–1954) investigated "finite combinatory processes" and anticipated many later results of other investigators, though he had trouble publishing some of them; his idea of a "production system" is based on that of Axel Thue (1863–1922). The lambda calculus, invented by Alonzo Church (1903–), is widely used in computer science for the description of programming language semantics (e.g. MacLennan, *FP*, Chs. 8, 9) and is the basis of the programming language Lisp. The basic ideas of combinatory logic were proposed by Moses Schönfinkel, but developed most fully by Haskell Curry (1900–1982); see Curry, Feys & Craig (*CLI*). Though begun by Peano and Thoralf Skolem (1887–1963), Jacques Herbrand (1908–1931) did much work on (general) recursive functions, which was continued by Kurt Gödel (1906–1978), and developed into a model of computation by S. C. Kleene (1909–); descriptions are widely available in books about the theory of computation. There are many accounts of the Turing machine model of computation, developed by Alan Turing (1912–1954); it stems from his work in mathematical logic, though he also applied it to the design of cryptographic computers during World War II. Markov algorithms were developed by A. A. Markov (1903–); they are described in Section 6.4.2. Markov algorithms have inspired several programming languages for string manipulation, most notably Snobol.

Turing computability, a proposal commonly known as *Church's Thesis*. It is not so much a statement of fact as a suggestion of a way of speaking.

Other Definitions of Computation

Despite the importance of Turing computability and its apparent correspondence with practical computability on modern digital computers, it's nevertheless important to keep in mind that it's just one model of computation, and that there are others both more and less powerful (see p. 414).

6.4.2 Markov's Normal Algorithms

> Altogether men have occupied themselves with mathematics quite a long time — no less than 4,000 years. During this period not a small number of different algorithms have been devised. Among them not a single non-normalizable one is known. After all this is a weighty argument in favor of the principle of normalization. No less weighty than, say, the experimental confirmation of the law of conservation of energy.
>
> — A. A. Markov (*TA*, p. 109)

String Substitution

Although the Turing machine is the best known model of computation, for many purposes it's better to use A. A. Markov's formalization of algorithms, which he called *normal algorithms*, but are commonly known as *Markov Algorithms*.[28] This is because Markov algorithms are based on the idea of a *production rule*, originated by Axel Thue (1863–1922; see Thue, 1914), but still widely used in a more developed form in the rule-based systems of cognitive science, linguistics and artificial intelligence. Markov algorithms are defined in terms of the most basic operation of formal manipulation, the substitution of one string of characters for another. A Markov algorithm comprises a number of *rules*, such as:

$$\texttt{1+1} \Rightarrow \texttt{2}$$

When this rule is applied to a string it causes the first (leftmost) occurrence of '1+1' to be replaced by '2'. For example, one application of the rule to the string

[28] The principal source for Markov algorithms is Markov (*TA*), with chapters 1 and 2 containing a detailed discussion of the assumptions and conventions upon which they are based. Mendelson (*IML*, Sect. 5.1) is briefer and more readable, but ignores the assumptions and conventions.

`3+1+1+5+1+1+1+1`

yields the string

`3+2+5+1+1+1+1`

Two more applications of the rule yield:

`3+2+5+2+2`

The principal assumption underlying the Markov model of computation is that all effective procedures may be reduced to simple substitutions of one string for another — think of a mathematician working on a blackboard, sometimes erasing, sometimes writing. The requirement that the string replaced exactly match the left-hand side of the rule ensure that all intuition has been replaced by judgements that the same characters occur in the same order. In the remainder of this section we'll define Markov algorithms more precisely, and try to show that it indeed captures the idea of effective calculation.

Assumptions and Conventions

The idea of character string substitution might seem simple enough, but there are a number of issues that must be addressed in order to give a precise definition of computation. More so than most early investigators of the theory of computation, Markov is careful to point out the assumptions, agreements and conventions underlying his model of computation (Markov, *TA*, Sect. 1.1–1.4, 2.1–2.3).

Tokens

For example, that stings comprise characters drawn from some alphabet seems straightforward enough, but we must agree about what shall be taken as characters. For example, consider the mathematical formula:

$$x \neq \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Before we can treat this formally, we must agree on its constituent characters. For example, is '$\pm$' a single character, or a composite of two, '+' and '$-$'? Is '$\neq$' a single character? Or two superimposed? Or three? What are the limits of analysis? Should we consider the entire a formula an arrangement of simple arcs and lines? It doesn't much matter how we answer these questions, so long as we realize that we must agree as to what are the primitive, atomic elements of formulas. We will call them *tokens*.

For our purposes in this chapter, it will be sufficient to define a token to be a *connected* printed mark, which corresponds in most cases to the intuitive idea of a letter or character. Notice however that this definition requires us to consider certain familiar characters to be composite ('i', 'j', ';', ':', '?', '!', ' " ', ' " '). (For simplicity we will use lowercase Greek letters and other symbols that are connected.)

Types

Second, we must agree about which differences between tokens are to be considered significant, and which not. For example, which of the following tokens represent the same character?

a a *a* **a** a a *a* A a

Sometimes mathematicians treat 'a' and '**a**' interchangeably, but other times they distinguish between them. Again, the specific answer is not so important as the recognition that there must be agreement about the kinds of differences that are considered significant, and the kinds that are not. How we answer these questions will determine how many *types* of tokens we have. For example, in the above display we could have nine tokens all of type *letter a*, or tokens of various types, as many as nine types.

Formality also demands that we avoid tokens whose type can only be determined in context. Consider this symbol:

,

Whether it is an apostrophe, a comma, or a quotation mark depends not on its shape but on its spatial relation to its context:

'It's an apostrophe, comma, or a quote.'

Interchangeability Within a Type

It is a fundamental property of formal systems that tokens of the same type are interchangeable. As in a chess game, where any white pawn is interchangeable with any other white pawn, so also in formal systems any 'a' is interchangeable with any other 'a' of the same type. This property is what gives formal systems their independence of specific material embodiment. All that matters is the "shape" of the characters, that is, the distinctive features that distinguish its type.

Infinite Producibility

A final crucial assumption about the atomic constituents of formulas is the *infinite producibility* of tokens belonging to a type. Thus there is no limit to the number of 'a's we can make, and the use of a rule such as

$$\texttt{1+1} \Rightarrow \texttt{2}$$

will never fail because we've "run out of '2's."

Source of the Infinite in Mathematics

This may seem no more than a convenience, but it is crucial for the formalization of mathematics, and all models of computation make some analogous (usually unstated) assumption. For example, if we are doing finitary arithmetic with numbers represented as strings of Pythagorean terms:

$$\bullet\bullet\bullet\bullet\bullet$$

then infinite producibility guarantees that we can always produce another '$\bullet$', and so count one more. The unlimited sequence of numerals is a direct consequence of the unlimited producibility of tokens. Likewise, when real numbers are approximated, it is infinite producibility that allows us to compute more decimal places or more elements of an approximating series. In the more sophisticated formalist approach the tokens do not represent numbers directly, but tokens are the raw material of proofs, which represent indirectly numbers, sets, and other infinitary objects. Without infinite producibility there would be no infinity in formalist mathematics.[29]

Potential vs. Actual Infinity

Though each type has associated with it an infinity of tokens belonging to that type, it is a *potential* infinity rather than an *actual* infinity. As Aristotle pointed out long ago, the distinction is crucial because it is the actually infinite that leads to paradoxes; the potentially infinite is comparatively untroublesome.

Linear Formulas

Mathematicians traditionally arrange their symbols in two dimensions, for example:

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This takes advantage of the capabilities of our visual perception and allows us to see the structure of the formula (MacLennan, ODFS). Nevertheless, formulas can always be written one-dimensionally, in a string, albeit with considerable loss of readability:

$$(-b \pm \sqrt{}[b \uparrow 2 - 4ac])/(2a).$$

For this reason, Markov restricted his attention to one-dimensional formulas, that is, to strings of characters.

Succession and Extent

By restricting our attention to linear formulas we avoid having to decide which of the following are superscripts, which subscripts, and which neither:

[29] Notice that infinite producibility is an *idealizing assumption*. If contemporary scientific opinion is correct, then the amount of matter-energy in the universe is finite, and so tokens made out of matter or energy are only finitely producible. The purpose of this note is

$$p^{n},\ p\,n,\ p\,n,\ p_{n}$$

but there are still potential problems, of which we should be aware. These lead to *distinctness requirements*. For example, when we write a string such as

to be or not to be

we must be clear about how much, if any, of the surrounding white space is included. We could adopt the convention that the string begins with the first nonblank character and ends with the last nonblank, but then we have a priori excluded strings that begin or end with blanks. A better solution, at least when confusion is possible, is to place the strings in quotation marks, so we can distinguish *'to be or not to be'* from *' to be or not to be '*.

Furthermore, if the amount of white space is significant, then we are obliged to use some notation such as '␣' to show explicitly the number of spaces. Thus we may distinguish '`2␣+␣1`' from '`2␣␣+1`' should we need to do so.

Uniqueness of Decomposition

Even though the tokens of different types can be distinguished in isolation, that does not mean that they can be distinguished when they occur in strings. For example, suppose that among our characters we have single and double quotation marks. Then the string:

’’’

could be parsed either as a single quote followed by a double quote, or a double followed by a single, or three single quotes. In other words, the string does not have a *unique decomposition*, which means that we cannot classify tokens unambiguously. (This problem does not arise if we require tokens to be connected shapes, as suggested above.)

Formulas Depend Only on Types

A consequence of the interchangeability of tokens of the same type is that the formal properties of a string depend only on the types of its constituents. For example, if we have agreed that 'a', '`a`' and 'a' are all tokens of the same type, then the strings '$a + 1$', 'a + 1' and '`a` + 1' are all interchangeable. In this way the string becomes a *formula*, embodying form independently of substance. Therefore, calculation allows us to operate on mathematical entities, such as numbers and proofs, by means of manipulation of concrete formulas embodying their abstract structure.

Idealizations

to caution against overenthusiastic application of the theory of computation to physical systems (e.g., claiming the universe is or isn't a Turing machine); it is only an approximate model of physical reality (as currently understood).

It's important to recognize that the ability to perfectly classify tokens as to type is an idealization of reality; no matter how reliable, physical machines are never perfect, nor is human perception. Mathematics can be viewed as the investigation of the consequences of the combination of the idealization of perfect classification — which Markov calls "the abstraction of identity" — with the idealization of infinite producibility — which he calls "the abstraction of potential realizability":

> The abstraction of potential realizability, just as the abstraction of identity, are quite necessary for mathematics. On the basis of these two abstractions is founded, in particular, the concept of natural number. (Markov, *TA*, p. 15)

Alphabet

A Markov algorithm operates upon strings of tokens of certain specified types called the *alphabet* of the algorithm. The alphabet is usually defined by exhibiting one token of each type in some arrangement that clearly separates the characters. For example, we may list tokens of each type, separated by commas and surrounded by curly braces,

$$\{\alpha, \beta, \bullet, \triangleleft\}$$

which is effective so long as the alphabet does not include commas or curly braces. It is important to observe that at the most basic level of definition we can only point. We cannot define a token's type in terms of its parts because, by agreement, it has no parts. In the end, we must show examples and trust that they will be sufficient to distinguish the types.[30]

Rules

A Markov algorithm is a finite list of *rules*, each comprising two strings, a *pattern* and a *replacement*, separated by an arrow for clarity. For example, the rule

$$\bullet \sim \bullet \ \Rightarrow \sim$$

replaces the leftmost occurrence of '$\bullet \sim \bullet$' by '$\sim$' (and does nothing if there are no occurrences of it). Like alphabets, rules are specified by showing examples: we exhibit an instance of the pattern and replacement strings and trust that the general rule will be understood in the context of our conventions (i.e., agreements about significant and insignificant differences).

Ordering of Rules

[30] If a machine is to use the formalism, then the types may be definable in terms of measurements of continuous quantities. For example, we might define a '0' to be 1.5 ± 0.1 volts and a '1' to be 0.5 ± 0.1 volts; voltages outside these ranges are not in the machine's alphabet.

Markov algorithms are executed as follows. At each step of the algorithm the rules are tried in order. That is, we try the first rule, applying it at the leftmost place where it matches. If it doesn't match anywhere, then the second rule is tried, and so forth until a rule is found that will match. If no rule matches, then the algorithm *terminates naturally.* If some rule does match, then the algorithm continues, and regardless of which rule was applied, the next step starts again with the first rule. There is another way that the algorithm can terminate. If the rule that matches is a *terminal rule*, which is indicated by the symbol '$\Rightarrow \cdot$' between its pattern and replacement strings, then that is the last rule executed and the algorithm is done.

Example: Unary Addition

To make this discussion clearer, we'll consider several simple Markov algorithms, beginning with several for arithmetic on *unary numerals*, that is, on numerals that represent the number n by n tokens of the same type.[31] For example, to add $3 + 2 + 4$, we will apply the algorithm to the string

$$\bullet\bullet\bullet+\bullet\bullet+\bullet\bullet\bullet\bullet$$

The intended result is '$\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet$', representing 9. The alphabet for this algorithm is $\{\bullet, +\}$. Obviously, the simplest unary addition algorithm simply eliminates the plus signs from the string, in effect pushing together the strings of tokens. This is accomplished the following one-rule algorithm:

$$+ \;\Rightarrow$$

It will replace the leftmost '+' by nothing, so long as there is a '+', and the algorithm terminates "naturally" when there's not one.

Example: Unary Difference

A slightly more complicated example is the *difference* of two unary numerals. This is actually the *symmetric* or *absolute* difference of the numbers, since it takes no account of the sign of the result:

$$5 \sim 3 \quad = \quad 3 \sim 5 \quad = \quad 2.$$

Our goal is an algorithm over the alphabet is $\{\bullet, \sim\}$ that takes, for example, '$\bullet\bullet\bullet\bullet\bullet \sim \bullet\bullet\bullet$' into '$\bullet\bullet$'. This is accomplished by two rules, one that

[31] We make the customary distinction between numbers and numerals: *number* refers to the abstract mathematical entities (the Platonic forms, if you like), whereas *numeral* refers to the physical representations of numbers in accord with some system (e.g., decimal, binary or Roman numerals). Thus '$\bullet\bullet\bullet\bullet$', '4', '100', 'IV' are *numerals* representing the *number* four.

subtracts one unit from each numeral so long as they both have one, and another that gets rid of the leftover difference sign:

$$\begin{aligned} \bullet \sim \bullet &\Rightarrow \sim \\ \sim &\Rightarrow \cdot \end{aligned}$$

When used to calculate $5 \sim 3$ the algorithm rewrites the string as follows:

$\bullet\bullet\bullet\bullet\bullet \sim \bullet\bullet\bullet$	initial string
$\bullet\bullet\bullet\bullet \sim \bullet\bullet$	by first rule
$\bullet\bullet\bullet \sim \bullet$	by first rule
$\bullet\bullet \sim$	by first rule
$\bullet\bullet$	by second rule

Example: Multiplication

For a more interesting example, consider multiplication of unary numerals, so that '$\bullet\bullet \times \bullet\bullet\bullet$' yields '$\bullet\bullet\bullet\bullet\bullet\bullet$'. This is accomplished in much the same way Pythagorean figures are multiplied, by making a copy of the lefthand number (e.g., '$\bullet\bullet$') for each unit in the right-hand number (e.g., '$\bullet\bullet\bullet$'). (Notice how infinite producibility is assumed.) However, a number of extra symbols are required to keep the intermediate stages in order. Thus the algorithm is defined over the alphabet $\{\bullet, \times, =, \triangleright, \triangleleft, \circ, \otimes\}$. The algorithm comprises ten rules:

$$\begin{array}{r|ccc}
1 & \bullet\triangleleft & \Rightarrow & \triangleleft\circ \\
2 & \triangleleft & \Rightarrow & \\
3 & \bullet\circ & \Rightarrow & \circ\bullet \\
4 & = \circ & \Rightarrow & \bullet = \bullet \\
5 & \otimes\bullet & \Rightarrow & \triangleleft\otimes \\
6 & \bullet\otimes & \Rightarrow & \otimes \\
7 & = \otimes & \Rightarrow \cdot & \\
8 & \triangleright\bullet & \Rightarrow & \bullet\triangleright \\
9 & \triangleright\times & \Rightarrow & \otimes \\
10 & & \Rightarrow & = \triangleright
\end{array}$$

There is not space here to explain this algorithm in detail, but the following will give some idea of how it operates, which is typical of Markov algorithms. (See also Fig. 6.2.) In the first step rule 10 puts '$= \triangleright$' on the left of the string. Then '$\triangleright$' scans right to the '$\times$' (rules 8, 9), which it turns into '$\otimes$'. This symbol deletes one from the right-hand number (rule 5), and initiates the duplication of the left-hand number to the left of the '$=$' sign. The

rule	string
initial	$\bullet\bullet\times\bullet\bullet\bullet$
10	$=\triangleright\bullet\bullet\times\bullet\bullet\bullet$
8	$=\bullet\triangleright\bullet\times\bullet\bullet\bullet$
8	$=\bullet\bullet\triangleright\times\bullet\bullet\bullet$
9	$=\bullet\bullet\otimes\bullet\bullet\bullet$
5	$=\bullet\bullet\triangleleft\otimes\bullet\bullet$
1	$=\bullet\triangleleft\circ\otimes\bullet\bullet$
1	$=\triangleleft\circ\circ\otimes\bullet\bullet$
2	$=\circ\circ\otimes\bullet\bullet$
4	$\bullet=\bullet\circ\otimes\bullet\bullet$
3	$\bullet=\circ\bullet\otimes\bullet\bullet$
4	$\bullet\bullet=\bullet\bullet\otimes\bullet\bullet$
5	$\bullet\bullet=\bullet\bullet\triangleleft\otimes\bullet$
1	$\bullet\bullet=\bullet\triangleleft\circ\otimes\bullet$
1	$\bullet\bullet=\triangleleft\circ\circ\otimes\bullet$
2	$\bullet\bullet=\circ\circ\otimes\bullet$
4	$\bullet\bullet\bullet=\bullet\circ\otimes\bullet$
3	$\bullet\bullet\bullet=\circ\bullet\otimes\bullet$
4	$\bullet\bullet\bullet\bullet=\bullet\bullet\otimes\bullet$
5	$\bullet\bullet\bullet\bullet=\bullet\bullet\triangleleft\otimes$
1	$\bullet\bullet\bullet\bullet=\bullet\triangleleft\circ\otimes$
1	$\bullet\bullet\bullet\bullet=\triangleleft\circ\circ\otimes$
2	$\bullet\bullet\bullet\bullet=\circ\circ\otimes$
4	$\bullet\bullet\bullet\bullet\bullet=\bullet\circ\otimes$
3	$\bullet\bullet\bullet\bullet\bullet=\circ\bullet\otimes$
4	$\bullet\bullet\bullet\bullet\bullet\bullet=\bullet\bullet\otimes$
6	$\bullet\bullet\bullet\bullet\bullet\bullet=\bullet\otimes$
6	$\bullet\bullet\bullet\bullet\bullet\bullet=\otimes$
7	$\bullet\bullet\bullet\bullet\bullet\bullet$

Figure 6.2: Example Execution of Markov Algorithm. The figure shows the successive states of the string resulting from an algorithm for multiplying unary numerals, in this case 2×3. See text for description.

$\triangleright 0$	$\Rightarrow$	$0\triangleright$	$\sigma\nu$	$\Rightarrow$	$\sigma 0\nu$
$\triangleright 1$	$\Rightarrow$	$1\triangleright$	$\sigma\kappa$	$\Rightarrow$	$\sigma 0\kappa$
$\triangleright$	$\Rightarrow$	$\triangleleft\nu$	$0\sigma 0$	$\Rightarrow$	00σ
$0\triangleleft$	$\Rightarrow$	$\triangleleft 0$	$0\sigma 1$	$\Rightarrow$	10σ
$1\triangleleft$	$\Rightarrow$	$\triangleleft 1$	$1\sigma 0$	$\Rightarrow$	01σ
$+\triangleleft$	$\Rightarrow$	$\oplus$	$1\sigma 1$	$\Rightarrow$	11σ
$0\oplus$	$\Rightarrow$	$+0\sigma$	0ν	$\Rightarrow$	0
$1\oplus$	$\Rightarrow$	$+1\sigma$	1ν	$\Rightarrow$	1
$0\sigma 0\nu$	$\Rightarrow$	$\triangleleft\nu 0$	0κ	$\Rightarrow$	1
$0\sigma 1\nu$	$\Rightarrow$	$\triangleleft\nu 1$	1κ	$\Rightarrow$	$\kappa 0$
$1\sigma 0\nu$	$\Rightarrow$	$\triangleleft\nu 1$	$\oplus\nu$	$\Rightarrow\cdot$	
$1\sigma 1\nu$	$\Rightarrow$	$\triangleleft\kappa 0$	$\oplus\kappa$	$\Rightarrow\cdot$	1
$0\sigma 0\kappa$	$\Rightarrow$	$\triangleleft\nu 1$	$\oplus$	$\Rightarrow\cdot$	
$0\sigma 1\kappa$	$\Rightarrow$	$\triangleleft\kappa 0$	$+$	$\Rightarrow$	$+\triangleright$
$1\sigma 0\kappa$	$\Rightarrow$	$\triangleleft\kappa 0$			
$1\sigma 1\kappa$	$\Rightarrow$	$\triangleleft\kappa 1$			

Figure 6.3: Markov Algorithm for Binary Addition. The alphabet is: $\{0, 1, +, \kappa, \nu, \sigma, \triangleright, \triangleleft, \oplus\}$. The symbols '$\kappa$' and '$\nu$' indicate carry or no-carry from a one-bit addition; the symbol 'σ' drags the rightmost remaining bit of the first number next to the rightmost remaining bit of the second number, and then triggers a one bit addition. The symbols '$\triangleright$' and '$\triangleleft$' are used for scanning to the right and left, respectively, and '$\oplus$' is a temporary substitute for '+' that causes 'σ' to begin its work. The algorithm, when applied to the string '`1001+111`' will yield the string '`10000`' after 29 steps; the reader may find it instructive to trace its execution.

duplication is accomplished by '$\triangleleft$' scanning left to convert the '$\bullet$' into '$\circ$' (rules 1, 2), which percolate to the left (rule 3) and convert to '$\bullet$' on both sides of the '=' sign (rule 4). When the multiplication is done, rules 6 and 7 eliminate temporary symbols no longer needed.

Binary Arithmetic

Although it's convenient to use unary notation for numbers, it is by no means necessary, and it's relatively straight-forward to design Markov algorithms to work in binary, decimal or any other base. For example, an algorithm for binary addition is shown in Fig. 6.3.

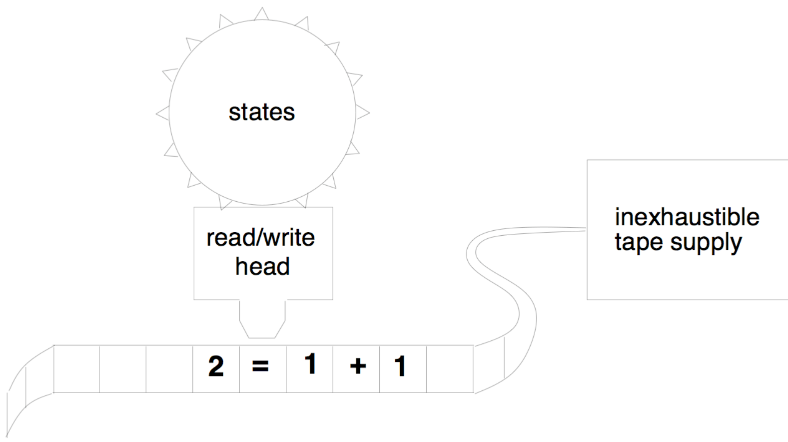


Figure 6.4: Depiction of Turing Machine.

Turing Machines

The Turing machine model of computation is similar to the Markov model in that it attempts to represent the purely mechanical reading and writing of symbols based on simple, syntactic decisions. In place of the string upon which the Markov algorithm operates is a tape, marked off into squares, each of which may bear a character from some specified finite alphabet (Fig 6.4). The picture is something like such early twentieth-century devices as punched paper tape readers and tickertape machines. The tape has a definite beginning, which we'll think of as the left end, upon which any input data is written. The tape may be extended at the right end as far as necessary to carry out the calculations. This "supply of tape that never runs out" is the Turing machine's principal source of the *infinite producibility* necessary for significant mathematics.

Corresponding to the finite rule set of a Markov algorithm, a Turing machine has a *finite state control.* At any given time the machine is in one of a finite number of states which determine how it will react to the symbols on the tape; the machine's current state can be thought of as a gear that can be in one of a finite number of definite positions (the possible states). Further, at a given time the machine's "read/write head" is positioned over exactly one square of the tape.

The behavior of the machine is determined by a combination of its state and the symbol under the head. In any such situation it can do exactly one of three things to the tape: (1) write a symbol to the tape, replacing whatever symbol is under the head, or (2) move the head one square to the right, or (3) move the head one square to the left. In addition, after the tape operation the machine can change its state — the gear can rotate to a new position. Thus the behavior of the Turing machine is completely specified by a (finite) table which, for each combination of state and tape symbol, specifies a new state and one of the three tape operations.

Computer scientists have investigated Turing machines with multiple heads, multiple tapes, and so forth. They are all computationally equivalent provided they are essentially finite, so the details of a Turing machine's architecture are not important so far as computability is concerned.

Markov algorithm simulation

It's relatively easy to show the equivalence of the Turing and Markov models of computation by a standard technique: showing that each Turing machine can be turned into an appropriate Markov algorithm (which shows Markov algorithms can compute everything Turing machines can compute), and showing that each Markov algorithm can be turned into a corresponding Turing machine (which shows that Turing machines can compute everything Markov algorithms can compute).

Suppose we have a Turing machine with states in Σ and tape symbols in T. We will define an equivalent Markov algorithm over strings in the alphabet $\Sigma \cup \mathrm{T} \cup \{|, \rangle\}$. If the tape contents are $\boxed{c_1 c_2 \cdots c_n}$, the state is s and the read head is positioned over the character $c_m = b$, then this situation will be represented in the Markov algorithm by a string of the form:

$$c_1 \cdots c_{m-1} |s\rangle c_m c_{m+1} \cdots c_n$$

Then for each state transition of the Turing machine we write Markov rules according to the following procedure.

If the machine, when scanning $c_m = b$ in state s, changes to state t, and rewrites b with a, then the new string will be:

$$c_1 \cdots c_{m-1} |s\rangle a c_{m+1} \cdots c_n$$

This is accomplished by the rule:

$$|s\rangle b \;\Rightarrow\; |t\rangle a$$

If instead, when scanning b in state s it moves to the right, then the new string is:

$$c_1 \cdots c_{m-1} b|s\rangle c_{m+1} \cdots c_n$$

This is accomplished by:

$$|s\rangle b \;\Rightarrow\; b|t\rangle$$

Conversely, if it moves to the left, then the new string is:

$$c_1 \cdots c_{m-2}|s\rangle c_{m-1} b c_{m+1} \cdots c_n$$

For this we use the rules of the form

$$x|s\rangle b \;\Rightarrow\; |t\rangle x b$$

for each $x \in \mathrm{T}$. In either case, if the new state t is a terminal state, then we use a terminal rule in the Markov algorithm. Finally, we add to the end of the algorithm a rule '$\Rightarrow\; |s\rangle$', which puts the TM head at the left end of the tape in the initial state s.

It's a little more complicated to convert a Markov algorithm into a corresponding Turing machine, but we'll summarize the method. The TM starts at the left end of the tape, and scans to the right to see if it matches the pattern on the left of the first rule. If it does, then it replaces the pattern by the right-hand side of the rule, perhaps moving the rest of the contents of the tape to the left or to the right, as necessary. If the first rule's pattern does not match the beginning of the tape, then the head is moved one position to the right, and an attempt is made to see if the first rule matches there. If the first rule matches nowhere on the tape, then the TM "rewinds" the tape and goes through the same procedure, but looking for the pattern on the left of the second rule, and so forth. Whenever a rule is successfully applied, the machine moves back into its original state, trying to apply the first rule to the beginning of the tape. Although the conversion is complicated and many details have been ignored, it should be apparent that the Markov to Turing translation is always possible.

6.5 General Assumptions of the Discrete

6.5.1 Invariants of Calculi

Goals

In this section I will summarize the properties of calculi, since they are the basis of both discrete knowledge representation and digital computation. In this way assumptions underlying calculi will be made explicit, and we will better understand the scope and limitations of results such as those of Turing, Gödel, Löwenheim and Skolem, which are discussed in Chapter 7.[32]

Invariants

We begin with three *invariants* that characterize all things called calculi; we could say they are the *essence* of calculi. These are *formality*, *finiteness* and *definiteness*. Each will be considered briefly.

Formality

Something is *formal* if it depends on *form* alone, which implies, in particular, (1) that it is independent of *substance* (i.e. abstract vs. concrete), and (2) that it is independent of meaning (i.e. syntactic vs. semantic). For example, when we do formal operations in arithmetic or algebra, on the one hand it doesn't matter whether the symbols are written with chalk on a blackboard or with pencil on paper; on the other, it doesn't matter what the symbols means, since we pay attention only to their shape and arrangement. We can call these *abstract formality* and *syntactic formality*.

Syntactic formality is important because it means that everything relevant has been captured in the symbol structures, for example, in the axioms of a mathematical system or in the data structures of an AI system or cognitive model. In this sense the system is purely *mechanical*; its operation does not depend on the presence of a meaning-apprehending intelligence. Abstract formality is important because it means that the formal system embodies all of the relevant properties, relations and processes; none of the mechanism is hidden in the physics underlying the system. In this sense, the formal system is completely *explicit*; everything it needs is in its formal structure.

Constructibility

Another invariant of calculi is *producibility*, which means that it's always possible to produce a text of the required form. Producibility is a corollary of abstract formality, since if something is independent of substance, its production cannot be limited by a lack of any substance. It means we take our model to be independent of any limitations peculiar to a particular physical realization.

Finiteness

[32] One of the best discussions of the underlying assumptions of discrete symbol systems is Goodman (*LA*, Ch. 5). It has been carefully considered in developing the present characterization, which is, however, somewhat different.

We have seen that from the time of Pythagoras, finiteness has been considered the sine qua non of intelligibility. First, formulas and knowledge structures are required to be finite in size (both extent and number of parts), otherwise we do not see how we can perceive them or manipulate them. For example, real numbers must be representable either as a finite ratio (rational numbers), or as a finite algorithm for computing its digits (computable irrationals), or in terms of a finite set of axioms determining its place in the real line (uncomputable reals). The Markov string and the Turing machine tape are always finite in length (though there is no limit to their growth).

Second, processes, such as inference and calculation, are also required to be finite in size (duration, number of steps, and complexity of steps), otherwise we cannot see how they could be completed. Each step of the Markov algorithm and each cycle of the Turing machine requires a finite amount of effort or energy. In both cases, a terminating process (one that gives an "answer") comprises a finite number of discrete steps.

Definiteness

Ancient Greek did not make a clear distinction between the finite and the definite; they constituted a single concept that we have written '(de)finite'. Thus, in their terms, a calculus is (de)finite and formal. This is a tidy way of expressing things, but for later discussions it's important to maintain the finite/definite distinction. We may say that something is finite if it has *external* bounds, but definite if it has *internal* bounds. So a calculus is bounded, both internally and externally.

Calculi have been taken to be *definite* since their first formulation, and definiteness is surely a major source of their attraction. For if any process is reduced to a calculus, then any complex or subtle judgements in the process have been replaced by simple and positive determinations. Recall Pythagoras' reduction of unintelligible expert acoustic judgements to the positive measurement of simple ratios. The ultimate motivation for definiteness is perhaps the desire for absolute certainty; as a culture we prefer yes-or-no answers to assessments of degree, and we prefer procedures that are easily recorded and executed, to practices involving skills painstakingly acquired through experience. And, in many cases, we prefer the predictable operation of a machine to the involved intervention of a human being.

In the following sections we will consider in more detail the properties of calculi, which will be codified in a series of *postulates*. It must be born in mind that these postulates are *assumed* to hold of any given calculus. There are probably no physical systems that satisfy them exactly, though some, such as digital computers, come quite close. A system can be considered a calculus to the extent to which it approximates this ideal.

6.5.2 Atomic Constituents

Definition (Token) *Tokens* are the atomic (indivisible) objects manipulated by a calculus.

Notice that tokens are atomic (indivisible) *by convention.* We cannot assume that they are physically indivisible; for example letters can be decomposed into lines and curves. The tokens are the smallest units manipulated by a given calculus; another calculus might manipulate their parts, and so work in smaller units.

Postulate Tokens are *definite* in their presence and extent.

That is, first we assume that with perfect reliability we can determine whether a token is present and separate it from the background. Second, we assume that we can tell when one token ends and another begins, that one cannot run into another in such a way that two tokens are indistinguishable from one.[33] This is a reasonable assumption for the paradigm cases of calculi. Think of the pebbles (literally *calculi*) of Pythagorean figured numbers, voting tokens, abacus beads, pieces for board games, printed words and formulas, holes in punched cards or punched paper tape, and electrical pulses in digital communication. That this postulate cannot be taken for granted is apparent when we consider less ideal examples, especially when calculi are used in noisy environments. For example, signal detection becomes a real problem when the signal-to-noise ratio is low, as in digital communication across deep space. Also a significant problem in reading a well-worn inscription or a faded manuscript is to determine whether a character is present and how many are there.[34]

Definition (Type) For the purposes of the calculus, certain tokens are mutually interchangeable; a group of mutually interchangeable tokens constitutes a *type*.

[33] It will be apparent that this postulate is fundamental to the formalization of whole-number arithmetic, since the distinguishability of the units is a necessary precondition to counting.

[34] The question of *whether* a token is present is logically prior to the question of *what* its type is, though in practice the two are often simultaneous. For an example in which they are not, consider the transcription of a text written in an unknown script.

Clearly, certain tokens may be interchangeable for the purposes of some calculi (and so constitute a type), but not for other calculi (as was discussed in connection with Markov algorithms, p. 249).

Postulate Types can be *definitely* determined.

That is, we assume the ideal ability to determine, without uncertainty or possibility of error, the type of a token. This is a good approximation to reality in the paradigm cases of calculi, such as digital computers. In borderline cases, such as those involving a low signal-to-noise ratio mentioned above, the postulate ignores the real-world problems of signal classification.

Postulate Types are *syntactically formal*.

Roughly, this means that a token's type is determined by physical properties, such as its shape or color, rather than by semantic properties, such as its denotation or connotation. However, this distinction reveals a dualist assumption, since it presupposes that semantic properties are not physical. If on the other hand we assume that the mind's ability to associate meaning with a symbol is a result of physical processes in the brain, then semantic properties are also physical.

To avoid building covert dualism into our definition of calculi (and thus begging important questions about the formality of cognitive processes), we must return to the pragmatic invariants that motivate the distinction between calculi and other things (MacLennan, CCKR, CSS). The importance of calculi rests on the reliability of its processes, which derive from simple, objective, nonproblematic discriminations, rather than from complex, subjective, arguable judgements or appeals to intuition. In this sense the syntactic formality of calculi is a corollary of the requirement that types be definite.[35]

Postulate Types are *finite* in number.

The underlying reason for this postulate is the assumption that an infinity of discriminations can be broken down into a finite number of more basic discriminations.[36] Due to the close mirroring of syntax and semantics in

[35]From a practical standpoint we can make a calculus with types based on semantic distinctions. For example, among English speakers, the bits 0 and 1 could be represented quite reliably by "common English words for animals" and "common English words for plants," even though this discrimination depends on the meanings of the symbols.

[36]For example, in treatments of formal logic one may read that there are a countably infinite number of variables, but there will always be an explanation that these can be written in a finite number of characters (for example, finite strings of letters of the alphabet).

calculi, this postulate is related to the assumption that every category can be defined in terms of a finite number of "essential" properties.

Postulate It is always possible to get or make a token of a given type (*producibility*).

Notice that producibility requires that there be a potentially infinite supply of tokens of any type. Therefore a type cannot be defined by simply exhibiting all the tokens of that type; rather, a type presupposes some general ability to classify tokens. In the case of an automatic calculus this ability is realized through some device, for example a circuit for distinguishing a "high" voltage pulse from a "low" one, or a short from a long pulse (as in Morse code). In cases where a human operates the calculus (as in the hand-arithmetic algorithms or a board game), we depend on the human's ability to recognize tokens of the different types.

6.5.3 Texts and Schemata

Definition (Text) A *text* is a group of interrelated tokens.

That is, a text is a composite physical object, which is manipulated by the calculus, and which typically represents knowledge, beliefs, hypotheses, etc. Traditionally, the tokens of a text were arranged spatially, in a plane (e.g., a sheet of paper or a blackboard) or in a line (Markov algorithm string, Turing machine tape). Tokens may also be related temporally as, for example, when electrical pulses are transmitted serially or words are spoken sequentially. The tokens stored in a digital computer's memory are related through the sequence of memory addresses. There are many other possibilities depending on the medium used to represent the text, that is, the background upon which the tokens are displayed.[37]

Postulate The extent of a text is *finite*.

That is, at any given time the text comprises a finite number of tokens, though that number may increase without bound. Loosely, the text has a beginning and an ending.

[37]For example, Curry (*OFPM*, pp. 28–29) gives an example of a calculus in which the tokens are sticks that are related by hanging from one another by strings, after the fashion of a mobile.

Postulate The extent of a text is *definite*.

That is, not only does the text begin and end, but we know (or can determine) where it begins and ends. The extent may be delimited either positively (e.g., by quotation marks) or negatively (e.g., by the absence of any tokens within a specified distance of the other tokens).

Definition (Schema) For the purposes of the calculus certain relationships among the tokens are considered equivalent; a group of equivalent relationships is called a *schema*.

For example, a certain range of spatial relations will constitute the schema "X is to the immediate left of Y." The classification of relationships into schemata is a matter of convention for a given calculus (as was discussed in connection with Markov algorithms, p. 250). Typically there must be some general mechanism — a device or a human — for classifying token relationships into the schemata of the calculus.

Postulate Schemata are *syntactically formal*.

That is, schemata are based on reliably determinable physical relationships, rather than on semantic properties (though the same remarks apply as previously made with regard to the syntactic formality of types).

Postulate Schemata are *definite*, though they may be ambiguous.

That is, we may definitely determine whether or not a particular arrangement of tokens belongs to a schema (For example, whether it represents an instance of "X to the immediate left of Y.") In principle, there is no reason to rule out ambiguity, though often calculi are designed either to be unambiguous, or so that ambiguity doesn't matter (as was done with Markov algorithms, p. 251). For example, we permit the same text 'XYZ' to represent two schemata: either X to the immediate left of YZ, or Z to the immediate right of XY. What is prevented by this postulate is the kind of fuzzy schemata membership illustrated by examples on page 250.

Postulate Schemata are *finite* in number.

This postulate precludes, for example, every possible orientation of a token being considered a different schema.[38] The justification is essentially the same as that for the requirement that there be a finite number of types.

[38]Such a scheme also violates the definiteness postulate, since the angles cannot be measured with perfect precision.

Postulate Schemata depend only on the types of the tokens.

That is, we cannot have a schema that holds for certain tokens of a type but not for others, since calculi presuppose that all the tokens of a type are interchangeable. For example, a white pawn being on white's square king-4 is a schema, whereas a particular pawn being in that position, or on a particular board, is not a schema.

Definition (Form) The *form* of a text refers to the types of its tokens and the schemata into which they are arranged.

Thus the form refers to exactly those properties of the text of possible relevance to the calculus; it is by virtue of these properties that the calculus is (abstractly) formal. Notice that the space in which texts are constructed will limit the possible combinations of types and schemata. For example, in a "normal" (simply connected) space for strings of characters, the same A-token cannot be both to the left and to the right of a single B-token. Since this definition of form presupposes a text embodying that form, our use of 'form' refers only to physically realizable combinations of types and schemata.

Postulate It is always possible to make a text of a given form (*producibility*).

Since the form is immaterial, we suppose that we can never run out of the resources needed to construct a corresponding text. (As noted above, the assumption of physical realizability is built into our definition of 'form'.)

6.5.4 Processes

Calculi have some value as static representations of information. Nevertheless, from their earliest uses (e.g. the abacus) their principal value has been their reduction of interesting processes — especially inference — to mechanical manipulation; this is most apparent in the modern digital computer. In this section we summarize the assumptions underlying the processes by which calculi alter texts through time. These processes include *algorithms*, the finitely and definitely specified procedures that form the central subject matter of computer science.

Definition (Process) A *process* alters a text through time, either by rearranging the tokens or by arranging new tokens.

For example, on a chess board or an abacus, the tokens are rearranged, but a person doing arithmetic or algebra on a blackboard typically rewrites the text at each step of the process.

Definition (Step, Operation) At each *step* of a process, an *operation* is applied to the text.

For example, in a board game a token may be moved, removed from the board, or placed on the board. In doing arithmetic by hand a digit might be written down or crossed out, or in using an abacus a bead might be moved.

Postulate Each step is completed at a *definite* time before the next step begins.

That is, the steps of a process form a discrete sequence. White finishes moving a chessman before black can start moving one. In doing long multiplication, we finish each one-digit addition or multiplication before we start the next. This postulate is an important precondition for definiteness, since it means that we know when each operation is done.

Postulate There are a *finite* number of possible operations in a process.

It's a fundamental principle that something can be specified precisely only if it can be indicated among a finite number of possibilities. Precise selection from an infinite set must be expressed as a finite number of selections from a finite set (e.g. a finite string of decimal digits).[39] Analogously, precise specification of the operation to be applied requires that there be a finite number of basic operations; more complex operations must be built from finite combinations of the basic operations.

Postulate Each operation is *local* (requires inspection of a *finite, definitely limited* part of the text, and similarly affects a *finite, definitely limited* part.

Nonlocal operations, which involve potentially unbounded regions of the text, likewise require potentially unbounded resources for their operation. Definiteness requires that holistic operations of this kind be broken down into more basic local operations that can be reliably specified and executed.

[39]If you have some doubt about this, think about the difficulty of using a notation with an infinite number of atomic symbols and no regular correspondence (such as in decimal notation) between the parts of a symbol and its meaning.

Postulate At each step it is *definite* which operations can be performed.

There are no doubtful cases or judgement calls. However, we allow the possibility of more than one operation being (definitely) applicable, since *nondeterministic* calculi are quite common.[40] For example, in proving a theorem in the propositional calculus there are typically several rules that can be applied at each step of the proof.

Postulate The applicability and result of an operation depend only on the *form* of the text.

That is, the steps are *formal*; both the material and the meaning of the tokens are irrelevant to the process. For example, we cannot have a game rule that allows me to capture this piece, but not another of the same type, and we cannot have a logic rule that does one thing with names of men and another thing with names of women (unless they are different in form).

Postulate The result of an operation is a text having a *definite form*.

Formality of course requires that only the form of the result is relevant. Furthermore, the form must be definite; we cannot have tokens of uncertain type or belonging to schemata that are not clear. In concrete terms, this means that the chess player must move the chessman to a specific square, the abacus bead must be definitely up or definitely down, etc.

Definition (Finite Process) A process is *finite* for a class of forms if it comprises a finite number of steps when operating on a text in any of those forms. It is called *total* if its finite for every form permitted by the calculus, and *partial* otherwise.

The importance of a finite process is that, since each step uses finite resources, so does the entire process. In particular, the process terminates in a finite amount of time, so these are the sort of processes from which we can expect an *answer* (final text) for any "question" (initial text) with a form in that class. We will see in the next chapter that there are important theoretical limitations on our ability to determine whether a process terminates over a given class.

[40] In a nondeterministic calculus situations arise in which more than one operation is applicable. In such situations some agent external to the calculus must determine which operation to apply. Examples of such agents include people (for manually operated calculi, such as games), physically random processes (roulette wheels) and deterministic processes external to the calculus (e.g., the rule stored last in the computer's memory).

Proposition A process can always be represented by a finite number of *rules*, each of which exhibits two texts, one exemplifying the form permitting an operation and one exemplifying the form resulting from the operation.

This proposition can be proved informally as follows. Since there are a finite number of types and a finite number of schemata, there are at most a finite number of distinct forms of any given size (number of tokens). Therefore, since operations are local, there are a finite number of forms that permit the operation; call these *conditions*. Furthermore, since the result of an operation depends only on the condition and serves only to alter the form of the text in a local region, we can specify for each condition the form of result of the operation for that condition; we call the resulting form the *effect* of the operation. Thus each operation can be specified by a finite number of condition/effect form pairs. Of course, we can't see a form, so each rule comprises a pair of texts exhibiting the forms. Finally, since a process has a finite number of operations, it can be specified by a finite number of rules. For a deductive system these rules are called the rules of inference; for a computer they are called a program.

The idea of a formal rule can be seen in its purest form in Markov algorithms, wherein the rules comprise pairs of strings. It is only slightly less apparent in the case of the Turing machine, as demonstrated by the corresponding Markov algorithm rules (p. 258).

6.5.5 Formal Semantics

Calculi are syntactically formal, so they don't "need" any meaning in order to operate. Nevertheless, *we* as users of the calculus normally attach some significance to the form of the text. Heretofore we have been concerned with *syntax*, the purely formal relationships among texts; now we turn to *semantics*, which addresses the relation between texts and their meanings. Semantics is a contentious subject, fraught with philosophical problems. Fortunately, for the purposes of this chapter and the next we will be able to limit our attention to *formal semantics*, a very specific sense of "meaning" that was developed by Alfred Tarski in the 1930s and works well for formalized mathematics (Tarski, CTFL; recall Sec. 6.2). In effect the meaning of texts is limited to mathematical domains, or to other domains defined with comparable precision. In Chapter **??** we will consider broader notions of meaning that are more appropriate for cognitive and psychological models. First, however, we need some terminology.

Although we speak of the meaning of a *text* and calculi operate by manipulating a physical text, abstract formality implies that only the *form* of the text is relevant. Therefore it is more correct to say that meaning is directly associated with a form and only indirectly associated with texts of that form. To avoid these awkward circumlocutions I will use:

Definition (Formula) A *formula* is a text that embodies a form relative to the type and schema conventions of a calculus.

Thus we can refer to the meaning of a formula and talk about the way a calculus transforms formulas.

Well-formed Formulas

Normally only a subset of the formulas of a calculus are meaningful; that is, some of the formulas are *well-formed* and others are *ill-formed*. For example, in a typical calculus of algebraic expressions, '$-b + \sqrt{b^2 - 4ac}$' would be a well-formed formula, but '${}^2 25\ 4 \times \times \sqrt{}$' is an ill-formed formula; a meaning may be associated with the first, but the second has no meaning, because it's gibberish.[41] It's quite possible that different purposes might oblige us to consider different classes of formulas to be well-formed.

Well-formed a definite property

Definiteness clearly requires that there be no doubt about whether a formula is meaningful, and since well-formedness is a precondition for meaningfulness, there likewise can be no doubt about whether a formula is well-formed. The following postulate is the usual way to guarantee that being well-formed is a definite property.

Postulate Well-formedness can be definitely determined by a total finite process.

That is, there is a process that, for each formula of the calculus, terminates in a finite number of steps and says whether or not that formula is well-formed. (This is called a *decision procedure* for well-formedness.) In computer terms,

[41] Of course we can imagine meanings for the second formula, and there is nothing that requires it to be ill-formed. The point is that normally we choose to restrict the class of well-formed formulas, since we are under no obligation to associate a meaning with the ill-formed formulas. Observe also that it's perfectly possible to have calculi in which every formula is well-formed, though they're usually rather simple calculi. For example, if our formulas are the strings of decimal digits, then every formula is well-formed as a decimal number.

there is a program that will determine whether any formula is well-formed.[42]

Mathematical Semantics

Tarski's mathematical semantics, as we've seen (Sec. 6.2), associates with each formula an *interpretation* or meaning in some *domain of interpretation*. For example, if the calculus is a formal axiomatic system for set theory (such as the common Zermelo-Fraenkel axioms, p. 298), then the "usual" (or "intended") domain of interpretation would be sets, and formulas would be interpreted as propositions about sets.[43]

Systematicity

Every calculus, except for the most trivial, permits an infinite number of formulas; therefore an arbitrary correspondence between formulas and their interpretations would require an infinite specification. This would contradict the *finiteness* requirements of calculi, so interpretations are required to be *systematic*, which means that by taking account of regularities in the correspondence, interpretations are finitely specifiable. The usual systematicity requirements are that the types and schemata, of which there are a finite number, each have a finite number of definite interpretations; thus the entire interpretation is finite and definite. This is expressed in the following three postulates.

Postulate Each type is interpreted as a *finite* number of *definite* objects in the domain of interpretation.

That is, each type corresponds to one or more mathematical objects (set, numbers, etc.) or to other precisely specified objects. This postulate permits ambiguous interpretations, in which a type may denote two or more objects, since calculi are often worth study. However, the ambiguity must be finite

[42] If the well-formed formulas are defined, as they usually are, by a formal context-free grammar, then there is a decision procedure for well-formedness: simply enumerate the (finite) list of well-formed formulas of the same size as the given formula, and see if it appears on that list.

[43] Tarski's theory of semantics makes most sense from a Platonic perspective, which is common among practicing mathematicians though not among philosophers of mathematics. Thus we imagine a Platonic realm of mathematical objects (such as sets) existing in itself, and we interpret the formulas of a calculus as propositions, which may be true or false about this realm. If instead we adopt a formalist philosophy of mathematics, then there is no realm of mathematical objects beyond the formulas themselves. In this case an interpretation reduces to a correspondence between two sets of formulas: those of the calculus and those of the formalized domain of interpretation; there is no semantics in the usual sense, only a syntactic correspondence between formulas. For this reason the notion of an interpretation is most comprehensible from a Platonic perspective.

and definite; there can be no haziness about which objects are denoted, and we must be able to specify them in a finite list.

Postulate Each schema is interpreted as a *finite* number of *definite* relations on the domain of interpretation.

For example, the schema of two formulas standing right next to each other might be interpreted as the product of the numbers denoted by the formulas. Again, there is no reason to prohibit systematic (finite, definite) ambiguity.

Postulate The interpretation of a formula is completely determined by its form.

That is, a formula's interpretation is given by nothing more than the interpretations of its constituent types and schemata; there are no context dependencies or other factors. For example, if the type '2' is interpreted as the number 2, the type 'π' is interpreted as the number π, and the juxtaposition schema is interpreted as multiplication, then the formula '2π' must be interpreted as the product of these numbers (two times π).

Summary

The foregoing postulates characterize every system that represents information in finite arrays of discrete symbols, from Pythagoras's arrangements of pebbles, to symbolic logic, to computer programs, to the most sophisticated (traditional) AI systems. The fundamental limitations discussed in the next chapter apply equally to them all.

Chapter 7

Limitations of the Discrete

> My purpose here has been . . . to show that a specific Gödel proposition — neither provable nor disprovable using the axioms and rules of the formal system under consideration — is clearly *seen*, using our insights into the meanings of the operations in question, to be a *true* proposition!
>
> — Roger Penrose (*ENM*, p. 116)

> The import of Goedel's conclusions is far-reaching, though it has not yet been fully fathomed. . . . Goedel's conclusions also have a bearing on the question whether calculating machines can be constructed which would be substitutes for a living mathematical intelligence. . . . There is no immediate prospect of replacing the human mind by robots. . . . None of this is to be construed, however, as an invitation to despair, or as an excuse for mystery mongering.
>
> — Ernest Nagel and James R. Newman (GP/WM)

Limits to the Reduction of Mathematics to the Discrete

In the preceding chapters we have discussed the 2500 year history of two related ideas, one epistemological, the other mathematical. The epistemological idea is that knowledge can be represented in the formulas of a calculus and that cognition is calculation — formal manipulation of those formulas. The mathematical idea is the arithmetization of geometry, which is motivated by the belief that the discrete is fundamentally more comprehensible than the continuous. The latter theme will be brought to its conclusion in this chapter, for we will consider several results that place fundamental limits on the arithmetization of geometry and on the axiomatization of mathematics. Although these results were established in the 1930s and were well-known to the scientific community by mid-century, philosophers, psychologists and AI researchers continued to use discrete, symbolic representations through most of the twentieth century. Therefore, the next two chapters will continue the historical presentation, and discuss the use of calculi in philosophy, cognitive science and artificial intelligence.

Reasons for Understanding the Proofs

In this chapter we will consider important theorems proved by Gödel, Turing and other logicians, and I will try to explain the proofs of these theorems. Nonmathematical readers may wonder why they are being subjected to these proofs, but the quotations that open this chapter show the reason. Gödel's theorem rivals quantum mechanics in the number of unwarranted conclusions it has engendered, often by mathematically sophisticated commentators. No doubt I am also misinterpreting the significance of these results, but I hope at least that readers who understand the proofs will be in a better position to draw their own informed conclusions about their significance. Nevertheless, some technical issues have been separated out, and I suggest that the remainder be skimmed if the going gets too tough. Be cognizant though of the risk you run by taking this route.[1]

Like a Rorshach test, quantum mechanics and Gödel's theorem invite the projection of our fears and hopes, and the popular fascination with these two ideas is perhaps a reflection of profound societal changes now in progress.

7.1 Undecidable Propositions

7.1.1 Gödel's Incompleteness Theorem

Kurt Gödel: 1906–1978

If an axiomatic system is consistent and complete, then for each proposition P, exactly one of the pair P and not-P is provable.[2] This is clearly the most

[1] Cognoscenti will no doubt be outraged by my informality, but I have tried to steer a middle course, avoiding a myriad of uninteresting details, while allowing a majority of readers to grasp the essence of the proofs.

[2] Gödel himself has provided a fairly readable, although somewhat oversimplified, overview of his proof in his original paper (Davis, *Undec.*, pp. 5–9). A well-known popular account is Nagel & Newman (*GP*), which is abbreviated in Nagel & Newman (GP/WM). See also "Gödel's Theorem" in Edwards (*EP*, Vol. 3, pp. 348–357). A good general reference for this chapter is Kneebone (*MLFM*), although there are many other discussions of

desirable situation, since then the axioms say neither too much nor too little. One of the landmarks of twentieth century logic is Kurt Gödel's 1931 proof that no "reasonably powerful" axiomatic system can be both consistent and complete. So that you will understand the significance of this theorem I will sketch its proof. (If you are interested in the details, see the appendices to this section, beginning on p. 284.)

Reasonably Powerful Axiomatic Systems

I've said that Gödel's result applies to "reasonably powerful axiomatic systems." What exactly does this mean? It will be most clear after we've completed the proof, for then you will be able to see what we've assumed. But I can give a rough definition now. We will make use of the usual laws of logic, including the law of the excluded middle. However, the proof is completely constructive, and appeals only to simple properties of the natural numbers. Thus it is acceptable even to intuitionists. Further we will assume that our axiomatic system is completely formal, so that the axioms are just strings of characters and the rules of inference are just string replacement rules (such as Markov algorithms). Since strings of characters can be encoded as integers (just think of the bit strings representing both), the resources of such an axiomatic system are adequate for talking about axiomatic systems (including itself). In the following, let $\mathcal{A}$ be any reasonably powerful axiomatic system.

To prove the completeness of an axiomatic system we must show that every proposition is decidable; to prove its incompleteness we must show that at least one is undecidable. Gödel's great accomplishment was to consider the possibility that axiomatizations of significant mathematics are incomplete, and thus to try to construct a counterexample.[3]

The Proposition Ω

Our task is: given a consistent axiomatic system $\mathcal{A}$, construct a proposition Ω guaranteed to be undecidable in $\mathcal{A}$. One way to accomplish this is to make Ω a proposition of $\mathcal{A}$ that asserts its own unprovability; then assuming the decidability of Ω will lead to a contradiction. For if Ω is provable then it's true, and hence unprovable (since Ω asserts its own unprovability). Conversely, if $\neg\Omega$ is provable (and hence true), then Ω must not be provable (since $\mathcal{A}$ is consistent), which means Ω is true (since it asserts Ω's unprovability). Again we have a contradiction. Thus we will have the incompleteness

---

these topics.

[3]Mathematicians had good reason to be optimistic about proving consistency and completeness; recall the consistency and completeness results discussed in Section 6.3. Von Neumann is reported to have reproached himself for not proving the incompleteness result because he had never seriously considered its possibility.

of $\mathcal{A}$ if such an Ω can be constructed.

If Ω asserts its own unprovability then it is a proposition about formulas in $\mathcal{A}$ and their derivability from the axioms of $\mathcal{A}$ by its rules of inference. Hence Ω is a proposition about *strings* and their relationships. Thus, if Ω is to be expressible in $\mathcal{A}$ then $\mathcal{A}$ must be able to express propositions about strings. Any reasonably powerful axiomatic system can do so; in fact it's sufficient that $\mathcal{A}$ be able to express propositions in elementary number theory (such as divisibility and prime numbers, see "Gödel Numbers," p. 284).

Indirect Self-Reference of Ω

Let ω be the string representing Ω; we've seen that the incompleteness of $\mathcal{A}$ will be established if Ω asserts its own unprovability:

$$\Omega \equiv \neg\mathsf{Provable}(\omega)$$

Unfortunately we have no guarantee such an ω exists. Indeed, since this equation looks suspiciously like Russell's paradoxical set (p. 226), we are well advised to question its existence and to seek a constructive definition. Our doubts are confirmed by attempting an explicit definition of Ω by replacing ω by '$\neg\mathsf{Provable}(\omega)$':

$$\begin{aligned}
\Omega &\equiv \neg\mathsf{Provable}(\omega) \\
&\equiv \neg\mathsf{Provable}(`\neg\mathsf{Provable}(\omega)\text{'}) \\
&\equiv \neg\mathsf{Provable}(`\neg\mathsf{Provable}(`\neg\mathsf{Provable}(\omega)\text{'})\text{'}) \\
&\ \ \vdots
\end{aligned}$$

Thus Ω looks like an infinite formula, which violates the finitary assumptions of formal systems. Therefore the construction of Ω must take a different tack.

The Diagonalization Argument

Suppose we make a list of all the decidable propositions of some form. If we can then construct a proposition of this form that is guaranteed to not be in the list, then we will have constructed an undecidable proposition. This suggests that we use a diagonalization proof such as Cantor used to show that for any list of rational numbers there is a real number that does not appear in that list (Section 6.1.3).

Therefore we will consider propositions of the form $P(s)$ where s is a string and P is a property of strings. We call such a property *decidable* if for each s the proposition $P(s)$ is decidable, and we call it *undecidable* if there is at least one s for which $P(s)$ is undecidable. Each property P is represented by a string p and each proposition $P(s)$ is represented by a string, which we write $\mathsf{subst}(p, s)$, which refers to the result of substituting s into p (see "Class Expressions," p. 285 for details).

Now consider all the strings $p_1, p_2, \ldots$ representing decidable properties of strings; let P_i be the corresponding properties. Since $P_i(p_j)$ is decidable for every i and j, we can make a table showing the truth or falsity of these propositions:

	p_1	p_2	p_3	p_4	p_5	p_6	$\cdots$
P_1	$\boxed{\text{T}}$	T	F	F	T	F	$\cdots$
P_2	F	$\boxed{\text{T}}$	T	F	F	T	$\cdots$
P_3	F	F	$\boxed{\text{F}}$	F	F	F	$\cdots$
P_4	F	T	F	$\boxed{\text{T}}$	F	T	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Since we assume the list includes *all* decidable properties of strings, we can construct an undecidable property by making it the negation of the diagonal (shown in boxes):

	p_1	p_2	p_3	p_4	$\cdots$
Q	F	F	T	F	$\cdots$

Thus we want $Q(p_i) \equiv \neg P_i(p_i)$.

Since Q is a property of strings we must define it in terms of the string p_i rather than the property P_i it represents; therefore we represent the proposition $P_i(p_i)$ by the string $\mathsf{subst}(p_i, p_i)$. Further, since the P_i are decidable properties, $P_i(p_i)$ is true just when $\mathsf{Provable}[\mathsf{subst}(p_i, p_i)]$ is true (see p. 285). These observations permit an explicit definition of the property Q:

$$Q(p) \equiv \neg\mathsf{Provable}[\mathsf{subst}(p, p)] \tag{7.1}$$

Clearly Q cannot appear in the list of decidable properties since $Q(p_i)$ is the negation of $P_i(p_i)$ for every i.

Since Q is an undecidable property there must be at least one string s for which $Q(s)$ is an undecidable proposition. Taking a clue from Russell's Paradox, we try q, the string representing Q. Then we have:

$$Q(q) \equiv \neg\mathsf{Provable}[\mathsf{subst}(q, q)] \tag{7.2}$$

This is exactly the string required, as we can see by letting $\Omega \equiv Q(q)$ and $\omega = \mathsf{subst}(q, q)$, which is the string representing Ω. Then Eq. 7.2 becomes

$$\Omega \equiv \neg\mathsf{Provable}(\omega) \tag{7.3}$$

Thus $\Omega \equiv Q(q)$ is exactly the undecidable proposition we sought. See Figures 7.1 and 7.2.

Summary of Gödel's Incompleteness Theorem

In an inconsistent axiomatic system every formula is provable. On the other hand, as we've just shown, in a consistent system that's sufficiently powerful (i.e., powerful enough to talk about strings or numbers) there is always a proposition Ω that's undecidable. (And note that we can actually construct this proposition; refer to "Definiteness of Ω," p. 286, to see it.) Thus, if such a system is consistent, it cannot be complete, and if it's complete it cannot be consistent. Alternately, no reasonably powerful axiomatic system can be both consistent and complete. This is *Gödel's Incompleteness Theorem*. Clearly, this result was devastating to the formalist program.[4]

7.1.2 Corollaries to Gödel's Theorem

Undecidability of Consistency

Gödel dealt an additional blow to the formalist program, for he showed that a (sufficiently powerful) consistent axiomatic system cannot prove its own consistency. This result is a simple corollary to the incompleteness proof. To see this, let C be any proposition asserting the consistency of $\mathcal{A}$; since everything is provable in an inconsistent system (p. 239), C can be an assertion that some well-formed formula is unprovable. This will do:

$$C \equiv \neg\mathsf{Provable}(\text{‘}0 = 1\text{’})$$

The proof of Gödel's Theorem can be easily formalized in $\mathcal{A}$, so we know that if $\mathcal{A}$ is consistent, then Ω is not provable in it. Since C implies the consistency of $\mathcal{A}$ we can prove the implication:

$$C \rightarrow \neg\mathsf{Provable}(\omega)$$

But $\Omega \equiv \neg\mathsf{Provable}(\omega)$, so we can likewise prove in $\mathcal{A}$ the implication:

$$C \rightarrow \Omega$$

Thus, if C (the consistency of $\mathcal{A}$) were provable in $\mathcal{A}$, then Ω would also be provable in $\mathcal{A}$. But since we've seen that Ω is not provable in $\mathcal{A}$, the consistency of $\mathcal{A}$ must be likewise unprovable.

---

[4] The proof outlined above requires $\mathcal{A}$ to satisfy a stronger property than simple consistency; it's called *ω-consistency*. I pass over this detail for three reasons: (1) reasonable axiomatic systems are ω-consistent; (2) the use of ω-consistency is buried in the proof of the correctness of $\mathsf{Provable}$ (p. 285), which I've omitted; and (3) Rosser (ESTGC) showed that Gödel's Theorem can be strengthened so as to require only consistency.

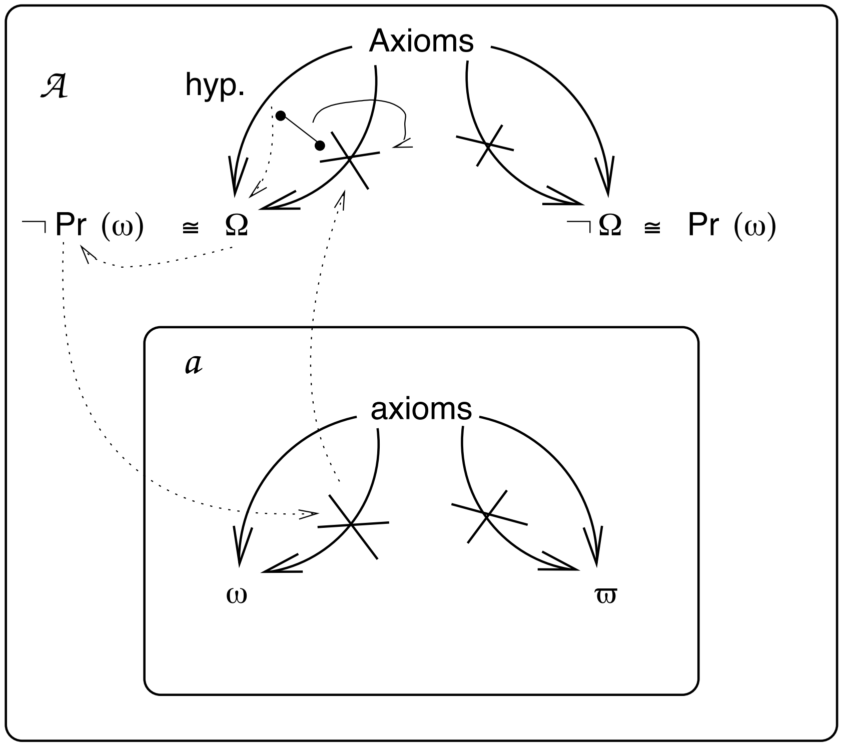


Figure 7.1: Gödel's Theorem, First Part. The diagram depicts the first part of the proof of Gödel's Incompleteness Theorem, showing that Ω is not provable. The outer box represents the axiomatic system $\mathcal{A}$, the inner box represents a, the axiomatic system encoded in Gödel numbers or in some other way that allows it to be a subject within $\mathcal{A}$. Dotted arrows indicate deductions from hypotheses later determined to be false; thin undotted arrows indicate true deductions. Lines with dots on both ends connect contradictory situations. Thick arrows indicate the possibility or impossibility of derivations from 'Axioms', the axioms of $\mathcal{A}$, or from 'axioms', the encoded axioms of $\mathcal{A}$ in a. We begin at 'hyp.' with the hypothesis that Ω is derivable from the axioms of $\mathcal{A}$. The dotted arrow shows that $\neg\mathsf{Provable}[\omega]$ is also provable. But the latter is provable if and only if ω is *not* derivable from the encoded axioms of a, as indicated by the crossed arrow in the inner box. Now derivations in a mirror those in $\mathcal{A}$, so we must conclude that Ω is not derivable in $\mathcal{A}$. This contradicts the hypothesis so we conclude that Ω is not provable in $\mathcal{A}$.

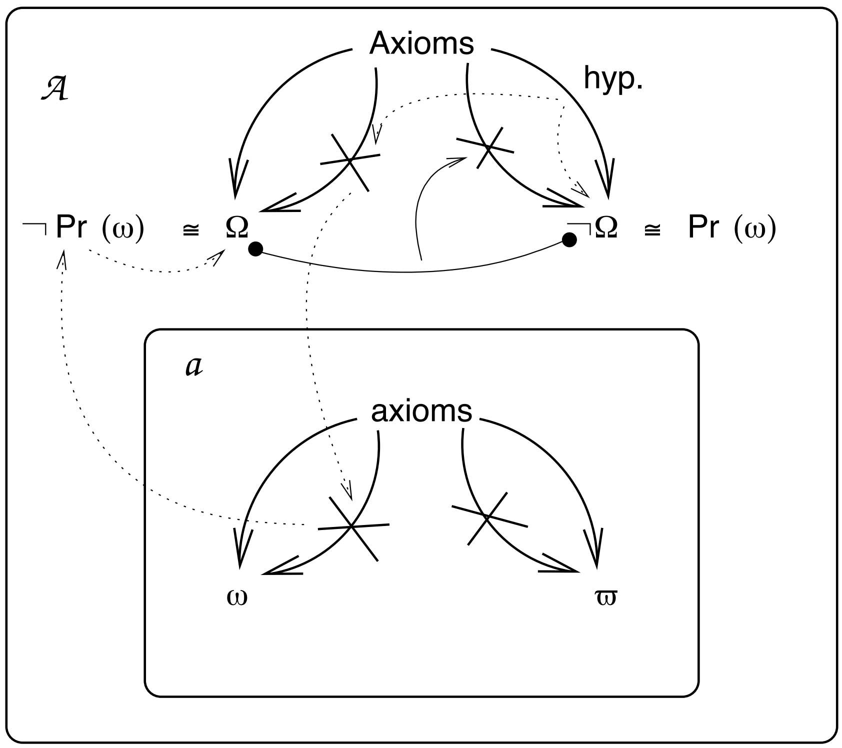


Figure 7.2: Gödel's Theorem, Second Part. The diagram depicts the second part of the proof of Gödel's Incompleteness Theorem, showing that $\neg\Omega$ is not provable in $\mathcal{A}$. Start with the hypothesis (marked 'hyp.') that $\neg\Omega$ is derivable from the axioms of $\mathcal{A}$, from which we conclude $\neg\Omega$. A dotted line indicates that the consistency of $\mathcal{A}$ allows us to conclude that Ω is not derivable. Therefore, in the encoded system a we know ω is likewise underivable. Hence we know $\neg\mathsf{Provable}(\omega)$, but this is exactly Ω, which contradicts $\neg\Omega$. Therefore we reject the hypothesis and conclude $\neg\Omega$ is not provable in $\mathcal{A}$.

In summary, Gödel showed that any reasonably powerful, consistent axiomatic system must have undecidable propositions, and that among these is the fact of its own consistency!

Metamathematical Provability of Ω

We turn to a surprising observation. We have seen that the formula Ω, which asserts the unprovability of Ω, is undecidable in the axiomatic system. Nevertheless, I claim that Ω is true, and prove it by the following *metamathematical* reasoning. We supposed that Ω is provable, and reached a contradiction. Therefore, applying the usual proof by contradiction, we must conclude that Ω is unprovable. That is, we have proved (metamathematically) that Ω is unprovable (in the axiomatic system). Since Ω asserts the unprovability of Ω in the axiomatic system [recall $\Omega \equiv \neg\mathsf{Provable}(\omega)$], we have proved Ω metamathematically. We've decided the undecidable proposition! Of course there's no contradiction here. We proved that Ω was undecidable in the given axiomatic system. It was this very fact that allowed us to then decide Ω by metamathematical reasoning — outside the system.

Although this is an important point, too much can be made of it. For example, the metamathematical proof has been the basis for claims that informal mathematics is inherently more powerful than formal mathematics (Penrose, *ENM*). Therefore the metamathematical proof deserves some scrutiny.

To many people the term *metamathematical* suggests some kind of supramathematical intuition, but, as we've seen, it simply denotes the use of mathematical techniques to reason about mathematics (see Section 6.1.4). This is exactly what we did in Gödel's proof when we defined predicates such as $\mathsf{Provable}$ and $\mathsf{IsaProof}$ (p. 284). Thus Gödel's proof is metamathematical. Also, contrary to some claims (Penrose, *ENM*), there is nothing inherently unformalizable about the metamathematical proof (see "Formalizing the Metamathematical Proof," p. 286). It is different from Gödel's proof in that it talks about the *truth* of propositions, whereas Gödel's talks only about their provability. Nevertheless, it's a routine exercise (see below) to construct a formal system $\mathcal{A}'$ capable of expressing propositions about the truth of the propositions of another system $\mathcal{A}$. Similarly an $\mathcal{A}''$ can be constructed that can express the semantics of $\mathcal{A}'$, and so on as necessary. It could be objected that this very argument shows the greater power of informal mathematics, since the informal metamathematical proof is valid for any axiomatic system $\mathcal{A}$, whereas the formal version requires constructing a new axiomatic system $\mathcal{A}'$ for each $\mathcal{A}$. Indeed, informal mathematics can talk about the truth of *its own* propositions. But even this self-descriptive ability may be formalized, since we can construct an axiomatic system $\mathcal{A}^*$ capable of expressing propositions about its own semantics. If we do so, however, we will make an interesting discovery: such an axiomatic system must be inconsistent since it is powerful enough to express a contradiction analogous to the Liar Paradox (p. 227): Define the predicate $Q(p) \equiv \neg P(p)$, where P is the interpretation of p, and consider the truth of $Q(q)$, where q is the encoded representation of Q.[5]

Again it might seem that the greater power of informal mathematics has been established, since a formal system with its expressive power must be inconsistent, but this does not follow. Since the Liar Paradox can also be expressed in informal mathematics, it follows that *informal mathematics is inconsistent*, just like $\mathcal{A}^*$. Indeed, the original Liar Paradox (p. 227) is a creature of informal logic, which is also inconsistent.

The phenomenon to be explained is not the *power* of informal reasoning, since it's already so powerful that it permits the Liar Paradox. Rather, the mystery to be solved is the process by which the community of mathematicians avoids perpetually encountering contradictions. It seems there must be *nonlogical constraints* that keep reasoning in check; I will address this issue in more detail in Section **??** (see also MacLennan, DD).

Implications of Gödel's Theorem

In the 60 years since Gödel published his result there has been little consensus about its implications. However, we can make the following observations. First, the result is extremely robust; it does not depend on details of the formal system. Obvious escapes, such as going to multivalued logics (logics with truth values in addition to true and false), do not change the result. There *are* systems ("semiformal" systems) that are complete and sufficiently powerful to prove their own consistency, but they diverge radically from the finitary goals of formalism. For example, some have infinitely large rules of inference, while others permit infinitely long proofs (Edwards, *EP*, Vol. 3, p. 355).

Certainly, if we restrict our attention to formal systems in the conventional sense, which presumes that they are finite (Section 6.5), then Gödel's theorem applies. Any such system (unless it's extraordinarily weak) must

[5]This assumes the axiomatic system assigns a truth value to every proposition and so also to $Q(q)$. It is of course possible to design a self-referential axiomatic system if it does not assign a truth value to propositions such as $Q(q)$. It also assumes $\mathcal{A}^*$ is powerful enough to talk about its own syntax (for which arithmetic is sufficient), and to talk about its own semantics (for which set theory is sufficient). See Beth (*FM*, pp. 335–345) for a detailed discussion.

have at least one undecidable proposition (unless the additional proposition made it inconsistent). And even if we add this proposition as an additional axiom, the resulting formal system must still have undecidable propositions. And yet all these propositions may be decided by metamathematical reasoning (which is just the garden variety mathematical reasoning applied to formal systems). Thus it seems that there is a sense in which a formal system can never capture the process of mathematics. This much is clear. Further implications are much less apparent. (See also Section 7.4.)

Gödel Numbers

Gödel wanted to reason about proofs, so he needed a representation for formulas and sequences of formulas. Now we would use strings of characters or linked lists, but Gödel didn't have these computer programming concepts, so he represented a sequence of numbers $n_1, n_2, \ldots, n_k$ by the number

$$N = p_1^{n_1} \times p_2^{n_2} \times \cdots \times p_k^{n_k}$$

where $p_1, \ldots, p_k$ are the first k prime numbers. By the prime factorization theorem, the i-th element of the sequence could be extracted by calculating the exponent of p_i in N. Sequences of characters were then represented by sequences of numbers, each number representing a character (now, we would probably use its ASCII code). Recall that Leibniz used the prime factorization theorem to represent finite sets (p. 118).

Provability

When we deal with an axiomatic system metamathematically, we treat formulas and proofs as string of characters (or, equivalently, natural numbers). Thus a relationship among formulas, such as being derivable by a given rule of inference, is just a relationship among strings (or natural numbers). Although it's tedious, it's not hard to define a predicate $\mathsf{Provable}$ so that $\mathsf{Provable}(p)$ means that the string p is derivable in the axiomatic system $\mathcal{A}$ from its axioms and by its rules of inference. Just to give the idea, here is the beginning of the top-down definition of this predicate:

$$\begin{aligned}
\mathsf{Provable}(e) &\equiv \exists p\{\mathsf{ProofOf}(p,e)\} \\
\mathsf{ProofOf}(p,e) &\equiv \mathsf{IsaProof}(p) \wedge e = \mathsf{last}(p) \\
\mathsf{IsaProof}(p) &\equiv \mathsf{Axiom}(p) \vee \\
&\qquad \exists q\, \exists s\{p = \mathsf{postfix}(q,s) \wedge \mathsf{DerivableFrom}(s,q)\}
\end{aligned}$$

These definitions make use of simple operations on sequences of strings (such as last, which returns the last element of the sequence, and $\mathsf{postfix}$, which adds an element to the end of the sequence), which also must be defined. Ultimately we get down to basic properties of strings (such as one being a substring of another), but these are easy to define in any reasonably powerful axiomatic system. If these definitions are carried out correctly, then we will be able to prove:

P is provable in $\mathcal{A}$ if and only if $\mathsf{Provable}(p)$ is provable in $\mathcal{A}$, where p is the string representing proposition P.

Class Expressions

By *formula* we mean a syntactically legal string in the language of the formal system $\mathcal{A}$, and by *class expression* we mean a formula with one free variable (i.e., one variable not "bound" by a quantifier). This is an example of a class expression:

$$\text{`}\exists m\{n = 2 \times m\}\text{'}$$

(In this case 'n' is a free variable and 'm' a bound variable.) Intuitively, this formula denotes the class of all even numbers.

Substitution for Free Variables

It is simple to write a program that substitutes one string for another. Therefore we assume that we have a function subst such that $\mathsf{subst}(p, s)$ replaces the free variable of the class expression p by the string s. For example, if $p =$'$\exists m\{n = 2 \times m\}$', then $\mathsf{subst}(p, \text{`17'})$ replaces 'n' by '17' yielding:

$$\mathsf{subst}(p, \text{`17'}) = \text{`}\exists m\{17 = 2 \times m\}\text{'}$$

As we've said, a class expression is intended to represent the class of numbers possessing the denoted property. Thus the formula returned by $\mathsf{subst}(p, s)$ can be interpreted as the proposition that s is a member of the class defined by p.

Definiteness of Ω

Notice that we have *constructed* the undecidable proposition Ω. To see this, recall

$$\begin{aligned} Q(q) &\equiv \neg\mathsf{Provable}[\mathsf{subst}(s,s)] \quad \text{and} \\ q &\equiv \text{‘}\neg\mathsf{Provable}[\mathsf{subst}(s,s)]\text{’} \end{aligned}$$

Then expand the definition of Ω:

$$\begin{aligned} \Omega &\equiv Q(q) \\ &\equiv \neg\mathsf{Provable}[\mathsf{subst}(q,q)] \\ &\equiv \neg\mathsf{Provable}[\mathsf{subst}(\text{‘}\neg\mathsf{Provable}[\mathsf{subst}(s,s)]\text{’}, \\ &\qquad \text{‘}\neg\mathsf{Provable}[\mathsf{subst}(s,s)]\text{’})] \\ &\equiv \neg\mathsf{Provable}(\text{‘}\neg\mathsf{Provable}[\mathsf{subst}(\text{‘}\neg\mathsf{Provable}[\mathsf{subst}(s,s)]\text{’}, \\ &\qquad \text{‘}\neg\mathsf{Provable}[\mathsf{subst}(s,s)]\text{’})]\text{’}) \end{aligned}$$

You can now see that Ω is a perfectly definite proposition; it and the corresponding string ω are 65 characters long (not counting blanks).

Formalizing the Metamathematical Proof

Requirements on $\mathcal{A}'$

To carry out a formal equivalent of the metamathematical proof would require many tedious constructions that would add little to understanding. Therefore my goal here will be to give just enough detail to make it plausible that the proof can be formalized. As before, we have the axiomatic system $\mathcal{A}$ and the undecidable proposition Ω constructed according to the Gödel procedure. Since $\Omega = \neg\mathsf{Provable}(\omega)$ *means* that ω, the embedded replica of Ω, is not provable in a, the embedded replica of $\mathcal{A}$, we see that Ω makes a true assertion. However, since the proof refers to the meaning of Ω, it's necessary to construct a model for $\mathcal{A}$. Therefore, the formal system $\mathcal{A}'$ in which the metamathematical proof will be expressed must be sufficiently powerful to allow the construction of formal interpretations. To accomplish this we need to be able to talk about the formulas of $\mathcal{A}$, for which arithmetic is sufficient, as we've seen, and we need to be able to define functions mapping these formulas into various subsets of the domain of interpretation, which is a set. Therefore, the mathematical apparatus of set theory is sufficient for defining interpretations, and set theory can be formalized by means of the Zermelo-Fraenkel axioms (p. 298) — though no one knows if they are consistent. Since ZF is sufficient to define arithmetic, we can take $\mathcal{A}'$ to be ZF without loss of generality.

Outline of the Proof

To show in $\mathcal{A}'$ that Ω is true, we must formally derive $\mathcal{I}\{\Omega\}$, the interpretation in $\mathcal{A}'$ of Ω. First express Gödel's proof formally in $\mathcal{A}'$; it should be clear that this can be done, because the proof uses only the most elementary proof techniques. Suppose the formal expression of the result is the following proposition of $\mathcal{A}'$:

$$\mathsf{Consistent}(\mathcal{A}) \rightarrow \neg\mathsf{ProvableIn}(\Omega, \mathcal{A}).$$

Now Gödel's proof hinges on the construction of the embedded system a so that '$\mathsf{Provable}(\omega)$' is derivable in $\mathcal{A}$ just when Ω is derivable in $\mathcal{A}$. Expressed formally in $\mathcal{A}'$ this is:

$$\mathsf{ProvableIn}(\Omega, \mathcal{A}) \leftrightarrow \mathsf{ProvableIn}(\text{`}\mathsf{Provable}(\omega)\text{'}, \mathcal{A}).$$

The interpretation of the latter proposition is:

$$\mathsf{ProvableIn}(\text{`}\mathsf{Provable}(\omega)\text{'}, \mathcal{A}) \leftrightarrow \mathsf{ProvableIn}(\omega, a).$$

Now notice that the interpretation in $\mathcal{A}'$ of Ω is:

$$\mathcal{I}\{\Omega\} \leftrightarrow \mathcal{I}\{\text{`}\neg\mathsf{Provable}(\omega)\text{'}\} \leftrightarrow \neg\mathsf{ProvableIn}(\omega, a).$$

Combining the implications we have:

$$\mathsf{Consistent}(\mathcal{A}) \rightarrow \mathcal{I}\{\Omega\}.$$

Therefore, we have a formal proof that if $\mathcal{A}$ is consistent then its Gödel proposition Ω is true. (Of course, an inconsistent axiomatic system has no models, and so we cannot even talk of its propositions being true or false.)

7.2 The Undecidable and the Uncomputable

7.2.1 Introduction

Alan Turing: 1912–1954

In this section we investigate Alan Turing's famous proof of the undecidability of the halting problem.[6] This result and its generalization — Rice's theorem — demonstrate inherent limitations to digital computation, and reveal an essential unpredictability in formal systems.

The Halting Problem

If you have ever programmed a computer you know that if you make a mistake your program may "go into an infinite loop." That is, it will run forever (or as long as you let it run), without ever stopping and returning an answer. A common predicament, when running a new program, is not knowing whether it's in an infinite loop. It's run for a minute so far, which is longer than you thought it should run. But does that mean that it's in an infinite loop, or only that it's slower than expected? You let it run another five minutes, and it still hasn't halted. Now you're becoming very suspicious, but you're still not sure that it won't return its answer in the next second or so. The trouble is of course that you never know for sure whether it will halt until in fact it does halts. It would surely be useful to have a way of telling in advance whether the program will halt. Then we would know we're not waiting in vain. This is the *halting problem*.

Decision Procedures

Since a program may halt on some inputs but not on others, we would like to know whether a given program will halt when run on a given input. A procedure (i.e., a program) for deciding this is called a *decision procedure* for the halting problem. We can imagine that this would be a very complicated procedure, analyzing the text of the program, and tracing its behavior on the given input. Nevertheless it would be valuable. There are of course many other questions we would like to ask about programs (when run on given inputs), such as whether they will ever try to divide by zero, whether they will return a particular output, and on and on. It would be useful to have decision procedures for all these problems. The remarkable thing that Turing proved is that there is no decision procedure for the halting problem, and a simple extension of his proof shows that there is no decision procedure for just about any property of interest. To understand this fundamental limitation of computers, it's important to see how it's proved. Therefore I'll

[6] The primary source for this section is Turing (OCN), which is reprinted in Davis (*Undec.*, pp. 116–154). Turing's proof is discussed in most books on computability theory and theoretical computer science.

present an informal overview of Turing's proof (using modern programming notations rather than Turing machines). The similarity to Gödel's proof will be apparent.

7.2.2 Undecidability of the Halting Problem

Assumed Decision Procedure

This will be a proof by contradiction, much like Gödel's proof. Therefore we suppose that we have a Boolean-valued procedure $\mathsf{Halts}(p, i)$ which returns **true** if program P halts on input i, and returns **false** otherwise. We assume that the program P is represented as a string of characters p in the obvious way.[7] Technically, p is a string representing a procedure declaration. For simplicity we will also assume that the input i is a string of characters; it will become clear that this does not limit the generality of the proof. In the Pascal[8] programming language the declaration of Halts would look like this:

```
procedure Halts (p, i: string): Boolean;
. . .
begin
. . .
end {Halts};
```

Altering the Diagonal

Turing's proof, like Cantor's and Gödel's, is a diagonalization argument. In this case, since we are considering programs whose inputs are strings, the diagonal is where the program is applied to itself (more precisely, to the string representing itself). When a program is "self applied" in this way it will either halt or not. As is usual in diagonal proofs, we will construct a

[7] I will use capital letters such as P to refer to programs (you can think of them as machine code loaded into the computer's memory). I will use small letters such as p to refer to the *source code* for the program: a string of characters in some programming language. Strictly speaking, a program P can be applied to some input, but program text p cannot. Also, a decision procedure can analyze program text p, but not the program P itself. For the most part these distinctions can be ignored, however.

[8] Pascal is one of the most popular modern programming languages. Although I use its notation, it should be clear that the proof could be carried through using any programming notation, including Markov algorithms. Turing of course used Turing machines (Section 6.4.2). The principal reference for Pascal is Jensen & Wirth (*PUMR*).

procedure that alters the diagonal. This procedure Q will halt if a given program does not halt when self applied, and will not halt, if the given program does halt under self-application. More precisely, $Q(p)$ halts if and only if $P(p)$ doesn't.

```
procedure Q (p: string);
. . . declaration of Halts . . .
begin
if Halts (p, p) then 1: goto 1;
end {Q};
```

This is how Q works. It takes the input string p (representing a program) and passes it to Halts as both the program and the input: Halts(p, p). We have assumed that Halts will tell us correctly whether $P(p)$ halts. If $P(p)$ does halt, then Q goes into an infinite loop (1: goto 1); otherwise it returns immediately (and therefore halts).

The Contradiction

It should now be obvious how we will get our contradiction. Let q be the program text representing the declaration of Q:

$$q = \text{`procedure Q (p: string); . . . end \{Q\};'}$$

Consider the result of applying Q to this string, $Q(q)$. As we saw, $Q(p)$ halts if and only if $P(p)$ doesn't halt. Therefore $Q(q)$ halts if and only if $Q(q)$ doesn't halt, which is a contradiction. More carefully, in executing $Q(q)$ we compute Halts(q, q). We have assumed that this tells us correctly whether $Q(q)$ halts. But whatever Halts says is contravened by Q; if it says $Q(q)$ halts, then Q loops forever; if it says it loops forever, then Q halts immediately. Thus our assumption, that Halts correctly decides the halting problem, must be wrong. We are forced to conclude that there is no decision procedure for the halting problem (see also "Picture of the Diagonal," p. 292).

7.2.3 General Undecidability

Rice's Theorem

We have seen that the halting problem is undecidable. You might suspect that this is a peculiarity of this problem, and that other interesting problems might be decidable. Unfortunately this is not the case. There is a generalization of Turing's results, known as Rice's theorem (Rice, CRES), which says that all interesting problems are undecidable. It will be easier to say what is meant by "interesting" after I sketch the proof.

etch of Proof

The proof follows the same outline as Turing's. Assume that we have a decision procedure DoesX(p, i), which tells us if a program P does something interesting X when applied to an input i. Then construct a diagonal procedure Q as before:

```
procedure Q (p: string);
. . . declaration of DoesX . . .
begin
if DoesX (p, p) then don't do X
else do X;
end {Q};
```

Discussion

In other words, if $P(p)$ does X, then $Q(p)$ doesn't do X; if $P(p)$ doesn't do X then $Q(p)$ does do it. The contradiction arises when we ask whether $Q(q)$ does X, for Q is constructed so that $Q(q)$ does X if and only if $Q(q)$ doesn't do X. Therefore there can be no decision procedure for determining whether a program does X. But what is X?

It is virtually anything. The only real restriction is that it must be in the power of the program to do it or not do it, otherwise we cannot construct Q. This includes just about any property of interest (e.g., dividing by zero, returning a particular number). Roughly, if it's not in the power of the language to do X, then there's not much point in a decision procedure that tells if a program does X, since in fact it never will.

Limitations on Undecidability Results

If we look carefully at the proof of these undecidability results, then we can see some hidden assumptions in them. Bringing these assumptions to light will help us to understand the scope and limitations of these results. We have already noted that the proof assumes that it's possible to "do" or "not do" the thing in question. In general, most logical properties of the program are controllable, although some physical properties (such as the amount of space or time used by the program) may not be. Another assumption is that the procedure Q can be constructed. For example, we have assumed that we can perform a conditional test (if . . . then . . . else . . .), although this is hardly a questionable assumption. More significantly, we have assumed that there is no limit on the size of a program. For example, if the largest program allowed were one million characters, and if it took 999 999 characters to define DoesX, then we would not be able to construct Q; it would be too big. Of course, when we define programming languages, and study the logical properties of computers, we avoid putting arbitrary limits on the sizes of things. On the other hand, it's important to keep in mind that most of these results depend on the potential infinities (i.e. infinite producibility, p. 249) that abound in the theory of computation and formal language theory. All real computers are finite, as are the programs that run on them. Real computers are equivalent to finite-state machines, not Turing Machines. Therefore we must be careful in applying these undecidability results to real computers and programs. (See the appendices beginning on p. 292 for the halting problem for finite state machines, and for an example of a property to which Rice's theorem does not apply.)

Conclusions

We have been talking about programs, but they are just the final culmination of the idea of a calculus: finite arrangements of uninterpreted tokens manipulated mechanically according to finite, formal rules. Thus these undecidability results inform us of the inherent limitations of discrete formal systems (calculi). On one hand, formal systems are too weak: they are incapable of deciding many interesting questions, in particular, most any property of formal systems in general. On the other hand, they are too powerful. They are so unpredictable that most of their interesting properties are undecidable by any rigorous (mechanizable) process. Formal systems are too weak to determine their own power.

Picture of the Diagonal

Let $p_1, p_2, \ldots$ be a list of all the procedure declarations and $P_1, P_2, \ldots$ the corresponding procedures. Then we can make a table of the truth value returned for each pair (P_i, p_j). The table might look like this:

	p_1	p_2	p_3	p_4	p_5	p_6	$\cdots$
P_1	$\boxed{\text{T}}$	T	F	F	T	F	$\cdots$
P_2	F	$\boxed{\text{T}}$	T	F	F	T	$\cdots$
P_3	F	F	$\boxed{\text{F}}$	F	F	F	$\cdots$
P_4	F	T	F	$\boxed{\text{T}}$	F	T	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

The diagonal elements are in boxes. We have constructed a procedure Q whose halting behavior differs from the diagonal:

	p_1	p_2	p_3	p_4	$\cdots$
Q	F	F	T	F	$\cdots$

Now we can see the contradiction directly. We have assumed that every program appears somewhere in the list $P_1, P_2, \ldots$ But by construction Q differs in its halting behavior from every program in the list. Therefore Q cannot appear in the list. Since the existence of the procedure Halts is the only questionable thing required for the definition of Q, it must be Halts that doesn't exist.

Example of a Property Not Covered

For an example of a property X not under the control of the procedure Q, consider "halts in ten seconds". Intuitively is seems like this property ought to be decidable: just run the program for ten seconds, and at the end of that time return true if the program has halted and false if it hasn't. And in fact our proof of the general undecidability result does not contradict this, for it's not necessarily in our power to "do X". For example, suppose DoesX requires at least ten seconds to run (as it would for our hypothesized decision procedure). Then it's no longer in Q's power to halt within ten seconds, since more than ten seconds have already elapsed. Of course this does not mean that "halts in ten seconds" *is* decidable; it only means that our proof does not show that it's undecidable. However, our intuitive argument suggests it is decidable.

Decidability of Halting Problem for FSMs

All real computers are finite state machines (all the memory cells, registers, etc. together can be in only a finite — though very large — number of states). But we can decide the halting problem for an N-state finite state machine as follows: Run the machine for $N+1$ cycles. If it has not halted by $N+1$ cycles, then it's in an infinite loop, since there are only N states, and whenever it returns to a previously visited state it must thereafter repeat the states that followed that state. (This assumes that the machine is deterministic, i.e., that its future action is determined by its present state.)

Although real computers are finite-state machines, computer science theory uses the Turing machine model, since the number of states is so large. (If a computer has one megabyte of memory, then the number of states is $2^{8\times 2^{20}} \approx 4 \times 10^{2\,525\,222}$, and this doesn't count internal registers or auxiliary memory, such as disks. In effect, it's presumed that this number is sufficiently large that it's a good approximation to infinity. Nevertheless, it's worth keeping in mind that real computers are not *in fact* equivalent to Turing machines. Infinite producibility is an idealizing assumption.

Relation of Turing's and Gödel's Theorems

Relation to Gödel's Theorem

It is not coincidental that Gödel's and Turing's proofs are so similar; they are really equivalent results. For example, you can see that Turing's theorem implies Gödel's as follows. It is straight-forward, though rather tedious, to define in an axiomatic system $\mathcal{A}$ a predicate $\mathsf{Halts}(p, i)$ that is true just when the procedure P (defined by string p) halts on input string i. (This assumes that $\mathcal{A}$ is sufficiently rich to express the semantics of a programming language, but this requires nothing beyond elementary number theory.) The proposition $\mathsf{Halts}(q, q)$ (where q is the string representing procedure Q) must be undecidable in $\mathcal{A}$, since otherwise we could solve the halting problem as follows: Program a procedure to enumerate in order of increasing length all the proofs in $\mathcal{A}$. If $\mathsf{Halts}(q, q)$ is decidable, then we must eventually enumerate a proof of either $\mathsf{Halts}(q, q)$ or $\neg\mathsf{Halts}(q, q)$. Whichever we enumerate first gives the solution to the halting problem, but since this is unsolvable, the proposition $\mathsf{Halts}(q, q)$ must be undecidable. Thus $\mathcal{A}$ is incomplete.

7.3 The Löwenheim-Skolem Theorem

7.3.1 Background

Leopold Löwenheim: 1878–1957

Thoralf Skolem: 1887–1963

Gödel's Theorem says that we can never adequately axiomatize a modestly rich body of knowledge, since our axioms must be either inconsistent or incomplete. This is an important limitation on the power of formal systems. Before we leave this topic, however, we must discuss another result, which limits them in a different but equally significant way. This is the Löwenheim-Skolem Theorem (1915, 1920), which implies that no axiomatic system can uniquely characterize the real numbers, or even the integers.[9] However, to state this theorem, we will need some terminology.

Definition of Categorical

In Section 6.2 we said that a *model* is an interpretation that results in the axioms and theorems being true of the domain. Thus, in trying to axiomatize any body of knowledge it's important that the intended interpretation be a model of the axioms, and — if we want our axioms to uniquely characterize that body — that that interpretation be the only model. Axiomatic systems with only one model are sometimes called *categorical* axiom systems. (See also "Uniqueness of Models," on p. 297.)

Satisfiability

There is of course no guarantee that an axiomatic system has a model. For example, as we would expect, inconsistent systems have no models, because their doing so would require contradictions to hold in fact. Technically, we say that an inconsistent system is not *satisfiable*. It is certainly not obvious that even consistent systems are in general satisfiable, but this is in fact the case, a result first proved by Gödel and sometimes known as his Completeness Theorem. To prove this result we construct ad hoc interpretations in which the symbols are essentially interpreted as themselves. Thus the domain of the interpretation is a set of formulas (symbol strings), and the function and relation symbols denote functions and relations on symbol strings. This is a technical trick that allows the theorem to be proved, but it also has important implications, to which we now turn.

7.3.2 The Theorem

Skolem's Parado

We've argued that any consistent axiomatic system has a model in which the domain is a set of formulas. But formulas are finite strings of characters from a finite alphabet, and so the set of formulas must be at most denumerable (see below, p. 298). Thus a consistent axiomatic system has a denumerable model. This is essentially the Löwenheim-Skolem Theorem.

Now this is a remarkable result. To see this, consider a typical axiomatization of set theory, such as the Zermelo-Fraenkel axioms, so long as expressed in finite texts over a finite alphabet. These axioms are often taken as the foundations of all mathematics (see p. 298). In particular, the ZF (Zermelo-Fraenkel) axioms are sufficiently powerful to prove the existence of the real numbers, and to express Cantor's diagonalization proof of the nondenumerability of the reals. Thus the ZF axioms assert the existence of a nondenumerable domain of objects. But the Löwenheim-Skolem Theorem tells us that these axioms (if they are consistent) must have a denumerable model. Thus, even though we can prove in this system that there's no one-to-one correspondence between the integers and the real numbers, there is some denumerable domain of objects, which includes all the objects that are called real numbers in the system. This is Skolem's Paradox: In our system we prove the theorem that the individuals are not denumerable. By definition, the theorems make true statements about the domains of the system's models. But we know that this system has a model with a denumerable domain. How can this be?

Resolution of Skolem's Paradox

The explanation of Skolem's Paradox seems to be this. The denumerable model of ZF set theory contains objects, functions and relations corresponding to the symbols of the axiom system. Suppose $\Re$ represents the set of real numbers in this system. Since the model is denumerable, there will be a denumerable number of objects in the domain that make the proposition '$x \in \Re$' true. Therefore, relative to the model, the reals are denumerable. On the other hand, we can prove in the system that there is no enumeration of all the x such that $x \in \Re$. But this means that there is no function in the domain capable of enumerating the objects that correspond to these x. Thus, although these objects are denumerable relative to the model, they

[9] Leopold Löwenheim: 1878–1957; Thoralf Skolem: 1887–1963. Discussions of the Löwenheim-Skolem Theorem can be found in Kneebone (*MLFM*, pp. 70–78, 303–307) and the article "Systems, Formal, and Models of Formal Systems" in Edwards (*EP*, Vol. 8, pp. 61–74).

are nondenumerable relative to the formal system. Although this explanation resolves the paradox, it leaves us without any absolute notion of denumerability. What had been well understood is now problematic.

Implications

It must be noted that Skolem's Paradox is not a flaw in some particular axiomatic system (such as Zermelo-Fraenkel set theory). It is a property of any consistent axiomatic system. From it we conclude that any attempt to axiomatize the real numbers (or any other nondenumerable domain) must fail, because the axioms permit as models essentially different (non-isomorphic), denumerable domains. We cannot have a categorical axiomatization of the real numbers (p. 295). In this sense the attempt to reduce the continuous to the discrete has failed. But the situation is worse than this, for a corollary of the Löwenheim-Skolem Theorem shows that a consistent axiomatic system must have models of all transfinite cardinalities (Edwards, *EP*, Vol. 8, p. 72). For example, an axiomatization of the *integers* must have models that are nondenumerable, and hence essentially different from the intended model. Thus we are denied even a categorical axiomatization of the integers!

Uniqueness of Models

It's easy to see that in a strict sense models cannot be unique. For example, the natural numbers $0, 1, 2, \ldots$ (with the usual interpretations) form a model for the Peano axioms, but so do the symbols $\square, \square\flat, \square\flat\flat, \ldots$ (Interpret '0' as the symbol $\square$ and interpret '$\mathsf{succ}(n)$' to mean 'append the symbol $\flat$ on the right end of the string n'.) Thus, the best that we can hope for is that all the models are *isomorphic* (i.e., there is a one-to-one relationship between the objects of the two domains that preserves the functions and relations on the domains). In this case we have the correspondences

$$0 \sim \square, \quad 1 \sim \square\flat, \quad 2 \sim \square\flat\flat, \quad \ldots$$

Also, corresponding to the successor function on the natural numbers we have the operation of appending $\flat$ on the end of a string. Thus when we say that an axiomatic system is categorical, or that it has one model, we will intend by this that all its models are equivalent "up to isomorphism."

Why a Set of Formulas Must be Denumerable

To see why a set of formulas must be denumerable, suppose we have an alphabet of N characters; then every string of characters can be thought of as a base N number. Therefore to every formula there will correspond a natural number, and so there can be at most as many formulas as there are natural numbers.

Notice that this argument depends on both the alphabet and the sizes of the formulas being finite. As we've seen (Sec. 6.5), these have been characteristics of formal systems since Pythagoras' time.

Zermelo-Fraenkel Axioms

The *Zermelo-Fraenkel axioms* refers to the most commonly accepted axiomatization of set theory. Zermelo (1871–1953) proposed seven axioms in 1908; these were revised and two additional axioms were added in 1922 by Fraenkel (1891–1965) and Skolem (1887–1963). These Zermelo-Fraenkel-Skolem axioms are the most commonly used, and a version of them is shown below. Another common axiomatization was developed from 1925–1954 by von Neumann (1903–1957) and Bernays (1888–?). See Beth (*FM*, pp. 381–398) for a discussion.

Axiom of Extensionality $\forall S \forall T[\forall x(x \in S \leftrightarrow x \in T) \rightarrow S = T]$. That is, if two sets have the same members, then they are the same set.

Axiom of Empty Set $\exists S \forall x[\neg(x \in S)]$. That is, there is a set with no members.

Axiom of Coupling $\forall x \forall y \exists S \forall z[z \in S \leftrightarrow (z = x \vee z = y)]$. That is, for any x and y, there is a set whose only members are x and y.

Axiom of Power Sets $\forall S \exists P \forall T(T \in P \leftrightarrow T \subseteq S)$. That is, for any set S there is a set (the power set of S) whose members are the subsets of S.

Axiom of Union $\forall S \exists U \forall x[x \in U \leftrightarrow \exists T(x \in T \wedge T \in S)]$. If S is a set of sets, then there is a set whose members are just the members of the members of S.

Axiom of Infinity $\exists S[\emptyset \in S \wedge \forall x(x \in S \rightarrow x \cup \{x\} \in S)]$. This guarantees the existence of at least one infinite set, corresponding to the natural numbers, that's constructed from nested empty sets:

$$\{\ \{\}, \{\{\}\}, \{\{\{\}\}\}, \dots\ \}\,.$$

Axiom of Regularity $\forall S \exists T[S = \emptyset \vee (T \in S \wedge \forall x[x \in S \rightarrow \neg(x \in T)])]$. This axiom prohibits $S \in S$ and "unfounded" sets with infinite descending chains of members, $S_1 \ni S_2 \ni S_3 \ni \cdots$.

Schema for Axioms of Replacement $\forall x \forall y \forall y'[(F(x,y) \wedge (x,y')) \rightarrow y = y'] \rightarrow \exists S \forall x[x \in S \leftrightarrow \exists w(w \in T \wedge F(w,x))]$, where $F(x,y)$ does not contain y', T or w. This is an *axiom schema*, that is, a pattern for generating an axiom for each suitable, expressible relation F. This says, roughly, that there is a set corresponding to any "reasonable" property expressible in the system. More precisely, it says that for each expressible function F and each set T there is a set $S = F[T]$ that is the image of T under F.

Axiom of Choice $\forall y \forall z[(y \in S \wedge z \in S \wedge y \neq z) \rightarrow \exists v \forall w(v \in y \wedge [w \notin y \vee w \notin z])] \rightarrow \exists u \forall y[y \in S \rightarrow \exists v \forall t(t = v \leftrightarrow [t \in u \wedge t \in y])]$. This says, roughly, that for any indexed set of nonempty sets, there is a function of the indices that "chooses" members of the sets. That is, if S_a is nonempty for each $a \in I$, then there is a function F on I such that $F(a) \in S_a$.

7.4 Epistemological Implications

The conclusion is inescapable that even for such a fixed, well defined body of mathematical propositions, *mathematical thinking is, and must remain, essentially creative.* To the writer's mind, this conclusion must inevitable result in at least a partial reversal of the entire axiomatic trend of the later nineteenth and early twentieth centuries,

> with a return to meaning and truth as being of the essence of mathematics.
>
> — Emil Post

7.4.1 Limitations of the Discrete

Applicable to all Formalizable Knowledge

It is important to realize that the results described in this chapter apply to any body of formalizable (i.e. verbalizable) knowledge, not merely to theories that are currently expressed as formal systems.[10] Thus these limitations apply to the very ideal to which scientific knowing has aspired, for we expect that the basic truths of a scientific theory should be expressible in a finite number of words, and we also expect that at each step in reasoning about this theory we need to consider only a finite number of words (and thus that the inferential processes are finitary). The result is that a reasonable scientific theory is formalizable, and therefore the Gödel and Löwenheim-Skolem results apply to it. In this section we consider briefly the implications of these results for scientific knowing.

Inevitable Incompleteness

Gödel's Incompleteness Theorem shows us that a formalizable body of knowledge must be incomplete. In other words, there must be some questions about the subject matter that the theory does not permit answering. Conversely, a reasonably rich, consistent, complete body of knowledge cannot be expressed in a finite number of words. Thus the complete understanding of any subject matter must take a very different form from what has traditionally been expected of scientific knowledge. In this sense we can never *say* all that there is to *know* about a subject. (Compare Socrates, Section 2.4.3.)

Subject Matter Cannot be Characterized

Even if we cannot say everything about a subject, it would seem that we ought at least to be able to uniquely characterize what it is we are talking about, but the Löwenheim-Skolem Theorem says that this is not so. For example, we cannot characterize the real continuum in a finite number of words, since any attempt must also apply to sets of objects that are essentially different from the reals (i.e., that are not isomorphic to the reals).

Failure of Arithmetization of Geometry

[10] More accurately, they apply only if the required formal system is "reasonably powerful" (p. 275) and consistent. Since it seems likely that any scientific theory must include multiplication and division, the hypothesis of "reasonable power" will be suppressed in the following discussion.

Although it appeared that the arithmetization of geometry — the reduction of the continuous to the discrete — had been successfully accomplished (Section 5.3), we now see that it hasn't. Finite words cannot exhaust the continuum. Eudoxus and Euclid eschewed the arithmetization of geometry and founded each science on its own axioms (p. 56), but we now see that even this cannot succeed.

Necessary Informality of Knowledge

The limitations of the discrete go beyond its inability to encompass the continuous. As noted (p. 297), a corollary of the Löwenheim-Skolem Theorem shows that even the *integers* cannot be uniquely characterized in a finite number of words. Also, Gödel's Incompleteness Theorem applies to the integers. Like the Pythagoreans 2500 years ago, we have discovered in number theory an element that is irreducibly *irrational*, ultimately *illogical*. (Recall the full meanings of *rational* and *logos*, Section 2.2.2.) Thus the most important limitation of the discrete is not the lack of a theory of the continuous. Rather it is the weakness of any body of knowledge that is in principle formalizable, for such knowledge must be both incomplete and incapable of characterizing its subject matter, except in the most trivial cases.[11]

7.4.2 Transcending the Discrete

Open Systems of Knowledge

There seem to be several possible ways of escaping from the limitations of the discrete. One comes from rejecting an assumption that we owe to Plato and Aristotle, the assumption that true knowledge can be expressed in a *closed deductive system*, that is, in a system in which all the truths derive from a finite number of explicitly stated axioms. What is the alternative? Since a formal system is defined by its axioms and rules of inference, we may allow either or both of these to be *open-ended*. I will briefly discuss the possibilities.

Open-ended Axioms

If the set of axioms is to be open-ended, then we must provide a *nondeductive* process for extending it. (The process must be nondeductive, since otherwise the new "axioms" are just theorems in a conventional axiomatic system.) Examples are such *ampliative* inferential processes as *induction* and *abduction*; these may lead to the invention or revision of axioms on the basis of observation (see "Definitions," p. 302.).

[11] In general, none of these limitations apply to calculi with a finite number of possible formulas, or to finite domains of interpretation. Nor do they apply to certain simple infinite systems, as we saw in Section 6.3. However we face these limitations in any system powerful enough to be mathematically interesting.

Definitions

Ampliative: "In [ampliative inference] the facts summed up in the conclusion are not among those stated in the premisses. ... These are the only inferences which increase our real knowledge, however useful the others may be." (Peirce, *CP*, 2.669–693)

Abduction: "Abduction consists in studying facts and devising a theory to explain them." (Peirce, *CP*, 5.145)

Induction: "Induction is the experimental testing of a theory." (Peirce, *CP*, 5.145)

On Peirce's notions of ampliative (or synthetic) inference, abduction (or hypothesis) and induction, see Buchler (*PWP*, Ch. 11), Goudge (*ToP*, Ch. 6) and Rescher (*PPS*, Ch. 3).

Ampliative inferential processes cannot be considered merely temporary measures necessary only until a science is completed. Rather, the *practice* of observation, invention and revision must be considered an integral part of the body of knowledge. The nature of such "practical knowledge" or skill is a central topic of Part III.

Open-ended Inference

Traditionally, it has been assumed that there is a fixed set of finitary inferential processes — the laws of logic. Therefore, one possible means of transcending the discrete is to allow the set of inferential processes to be open-ended, to recognize situations in which the inferential resources of a formal system can be extended. Notice, however, that there cannot be formal rules for this extension, otherwise they could be made a part of a formal system, and the usual limitations would apply. This means that we cannot expect precise specifications of when an extension is allowed or of the results of the extension.

How then does inferential extension take place? It seems that there must be some process for proposing possible extensions and for judging if they should be made (else we will have no confidence in the resulting formal systems). There are several possibilities. One is that we may discover new, intuitively valid inferential principles that can be formalized as deductive rules. However, in the 2300 years since Aristotle first formalized logic, this has rarely happened, so it seems unlikely that this process will be a continuing source of new inference rules. (Remember: our goal is not to fill in the gaps in an otherwise complete theory (which is impossible), but to find processes that are continually productive of new rules of inference.)

A more likely possibility is that new rules will be identified and justified by their consequences. For example, new deductive rules may be accepted because they allow the derivation of classes of theorems for which we have empirical or other informal reasons to want to be provable. This is, necessarily, an informal, ampliative process. As in the case of open-ended axioms, we see that a body of knowledge with open-ended inferential processes must include a set of ampliative practices that cannot be expressed as formal rules.

Requirements for Completeness

Is there any sense in which an open body of knowledge could be considered complete? For this to be the case we would have to believe that the ampliative processes are adequate to answer any question askable in the system. Are there such processes? Many people believe that the empirical methods of the sciences are complete in this way, but the claim needs justification.

More generally we can ask, "What would be the nature of ampliative processes that *could* be complete?" In effect we want processes that are guaranteed to gives answers but that satisfy certain criteria of objectivity (e.g., public accessibility, replicability, criticizability). These are essentially social criteria, and their roles in a theory of knowledge are discussed later (Sections 11.4.4, 11.4.4 and 11.4.4).

Infinitary Knowledge

There is a more radical way by which we may transcend the discrete. As noted previously (p. 282), there are "semiformal" systems that are both consistent and complete, but they diverge radically from the finitary assumptions that underlie the traditional view of scientific knowledge. We can entertain systems that are *infinitary* in either their axioms or rules of inference (or both).

Infinitary Axioms

Since the time of Pythagoras it has been assumed that the basic truths of a science must be (de)finite (i.e., finite and definite). One way to escape the limitations of the discrete is to reject this assumption by permitting axioms that are *essentially infinite*, either in number or structure.[12] For

[12]By "essentially infinite" I mean that they cannot be generated by some regular (i.e.

Open-ended Mathematics

The need for nondeductive processes for extending the basic truths of a body of knowledge may not seem surprising for those sciences traditionally considered empirical (physics, biology, psychology, etc.). However the Gödel and Löwenheim-Skolem results show that ampliative inference is also necessary in mathematics. Although this is not widely acknowledged, the history of mathematics exhibits many nondeductive processes; see Section 11.4.4.

example certain propositions about all the points of a continuum cannot be expressed in a finite number of discrete symbols, although they could be finitely expressed in a *continuous language* (for example a language whose "formulas" are images drawn from a continuum). Further, if the axioms are themselves drawn from a continuum, then there will be axioms that are arbitrarily "close" to one another; in this sense we may call them *indefinite* axioms (Fig. 7.3).

Infinitary Inference

There are two senses in which inference may be infinitary: in the rules of inference or in proofs. Traditionally, semiformal systems permit rules of inference with an infinite number of premises, such as the "rule of infinite induction" (Edwards, *EP*, Vol. 3, p. 355). For example, infinitary rules of inference could embody continua in their antecedent; it would be natural to have continuous inferential rules of this kind to go with the continuous axioms described above. Such rules could be finitely specifiable in a continuous language (as would continuous axioms), but the decision as to whether a rule is applicable might require arbitrarily precise discriminations.

Another source of infinitary inference is to allow infinite proofs. These could of course be proofs with an infinite number of discrete steps of the usual kind, but finite proof length can be preserved in a system with *continuous* proofs, in which the theorems evolve from the axioms by a continuous process. (Think of classical mechanics to picture this possibility.)

Continuous Knowledge Representation

What are the implications of infinitary systems for knowledge representation? Here it will be helpful to distinguish the two senses of the (de)finite:

finitely specifiable) way.

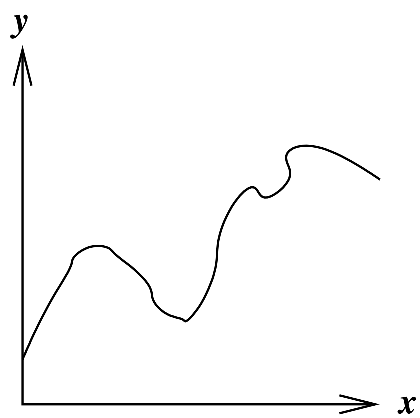


Figure 7.3: Simple Example of a Continuous Rule. Just as an discrete axiom '$P \vee P \Rightarrow P$' shows how to map the formulas of one discrete space into those of another (for example, this one takes '$A \vee A$' into 'A', '$(p \wedge q) \vee (p \wedge q)$' into '$(p \wedge q)$'), so the *graph* above shows how each point in one continuum (represented by the x axis) can be mapped into a point in another continuum (the y axis). Note that the graph (the curved line) itself is the finite, written representation of the correspondence, just as the string '$P \vee P \Rightarrow P$' is in the discrete case. Note also that there are infinitely many graphs arbitrarily similar to the graph shown above.

the definite (or discrete) and the finite (or bounded), for while we've found the limitations of the discrete, there are still advantages to hewing to the bounded. The reason is that if we are interested in the representation of knowledge in people and computers, then we must limit ourselves to representations that are physically realizable, which means that they must be bounded (require finite matter, energy and time).

Hence, it seems that we may escape the limitations of calculi while saving their physical realizability by representing knowledge in continuous, bounded structures — what topologists call *continua*.[13] When analyzed independently of its physical embodiment, such a structure may be called a *continuous formal system* or a *formal continuum*, the principal topic of Part IV.[14]

[13]This term is defined in slightly different ways by different authors; I will take a continuum to be a connected compact metric space with more than one point (Iyanaga & Kawada, *EDM*, §81C).

[14]MacLennan (LNAI) argues for the necessity of "continuous logics" and presents two examples. Both are based on continuous "propositions" and continuous rules of inference, but one has the traditional discrete derivations while the other uses continuous derivation. Some steps toward continuous formal systems can be found in MacLennan (CCKR, CSS, GAC, WLIOW, IS). Already a number of theoretical results show the ability of continuous computational models to transcend the limits of Turing computability (Pour-El & Richards, 1979, 1981, 1982; Stannett, 1990).

Chapter 8

Logical Positivism

> I want neither that plutocracy grasping and mean, nor that democracy goody and mediocre, occupied solely in turning the other cheek, where would dwell sages without curiosity, who, shunning excess, would not die of disease, but would surely die of ennui.
>
> — Poincaré (quoted in Runes, *ToP*, p. 966)

> Science itself, therefore, may be regarded as a minimal problem, consisting of the completest possible presentment of facts with the *least possible expenditure of thought*.
>
> — Mach (Newman, *WM*, Vol. 3, p. 1792)

8.1 Historical Background

Logical Analysis

In this chapter we will look at *logical positivism*, the most influential philosophy of science in the twentieth century. In spite of the fact that logical positivism has been abandoned by most philosophers of science, its influence continues in many disciplines, including physics, linguistics and psychology. We will be especially concerned with logical positivism's view of knowledge, which is, roughly: (1) the only real knowledge is scientific knowledge; (2) by a process of logical analysis scientific knowledge can be reduced to symbolic formulas constructed from "atomic facts." Certainly assertion (1) is nothing new; Socrates said as much when he distinguished "scientific knowledge" (*episteme*) from a "practice" (*empeiria*); see Section 2.4.3. Furthermore, assertion (2) is implicit in Pythagorianism and is a continuous theme in most Western epistemology, from Plato and Aristotle, through Hobbes and Leibnitz, to Boole and Hilbert. In this sense logical positivism is just the continuation of this long tradition.

Empiricism

The major innovation of logical positivism follows from a new understanding of scientific knowledge, and in the resulting view of the atomic constituents of knowledge. Throughout most of these 2500 years scientific knowledge was viewed rationalistically, and it was assumed that the atomic constituents were some kind of self-evident axioms involving basic categories that required no definition. However, beginning in the Renaissance there was a growing recognition of the value of observation and experiment (active intervention in nature) and a corresponding increase in scepticism about the "self evidence" of any proposition. Indeed, if empiricist philosophers were willing to grant self-evidence to anything, it was to sense data, rather than to the metaphysical propositions favored by the rationalists. In the past the ideal of knowledge had been Euclidean geometry, but the new empiricism took physics as its ideal.

Anglo-American Roots

Although logical positivism originated in Germany in the 1920s, many of its roots are Anglo-American. First there is *empiricism*, a strong theme in British philosophy back to David Hume (1711–1776), Thomas Hobbes (1588–1679) and Francis Bacon (1561–1626). Second, there is American *pragmatism*, originated by C. S. Peirce (1839–1914), William James (1842–1910) and John Dewey (1859–1962), with its emphasis on the observable consequences of actions. Finally, there was *logical analysis*, strongly represented in the Continental idealist tradition, but turned into a technical tool by the British philosophers G. E. Moore (1873–1958) and Bertrand Russell (1872–1970).

Continental Roots

Auguste Comte: 1798–1857

Logical positivism combined these Anglo-American developments with the existing positivist tradition on the Continent, which originated with the French philosopher Auguste Comte. He argued that each branch of human understanding went through three stages: in the *theological* stage, explanations are based on the volition of gods; in the *metaphysical* stage, explanations are based on abstract forces; in the *positive* (or scientific) stage, phenomena are not explained at all, but are simply connected by laws. Comte says,

> Finally, in the Positive state, the human mind, recognizing the impossibility of obtaining absolute truth, gives up the search after the origin and destination of the universe and a knowledge of the

> final causes of phenomena. It only endeavors now to discover, by a well-combined use of reasoning and observation, the actual *laws* of phenomena — that is to say, their invariable relations of succession and likeness. (quoted in Runes, *ToP*, p. 262)

Jules Henri Poincaré: 1854–1912

The famous French mathematician and philosopher Jules Henri Poincaré applied positivist ideas to physics and was concerned with the construction of basic physical concepts out of our sensations and perceptions. In this way he prepared the way for Einstein's reconstruction of space, time and gravity in his special and general relativity theories (see p. 327).

Ernst Mach: 1838–1916

The Austrian physicist Ernst Mach took a very similar view, and argued that the laws of physics, as of all the sciences, are just shorthand summaries of relationships between the observer's experiences:[1]

> It is the object of science to replace, or *save*, experiences, by the reproduction and anticipation of facts in thought. ... This economical office of science, which fills its whole life, is apparent at first glance; and with its full recognition all mysticism in science disappears.

Every scientific law and concept — mass, energy, atoms — must be reduced to "relations between observable quantities." These in turn are relationships among sense data.

> Properly speaking the world is not composed of "things" as its elements, but of colors, tones, pressures, spaces, times, in short what we ordinarily call individual sensations.

This view led Mach to detailed investigations of sensation and perception; for example in his *Space and Geometry* (1906) he carefully distinguishes the space of physics from the physiological space upon which it's based, and in the latter category he analyzes the space of vision and the space of touch (which, of course, are not the same). These analyses formed a starting point for Einstein's relativity work (p. 327).

Rejection of Metaphysics

Much like Comte, Mach saw science as an advance over metaphysics. "Where neither confirmation nor refutation is possible, science is not concerned." He rejected anything not reducible to experience as not true knowledge — a characteristic attitude of positivism — and by reducing statements

[1] A relevant selection from Mach's *Science of Mechanics* (1883), from which these quotations are taken, is reprinted in Newman (*WM*, Vol. 3, pp. 1787–1795).

to experience, he hoped "to expose the real significance of the matter, and get rid of metaphysical obscurities." As Hume said long before,

> If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does is contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning, concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion. (*Enq. Conc. Human Underst.*, sec. 12, pt. 3)

In 1895 Mach became a professor of physics at the University of Vienna, and with him we come to the doorstep of the Vienna Circle, the fountainhead of the logical positivist movement. However, before we discuss it, we must consider important contributions from the analytic tradition, upon which the logical positivists also depended.

8.2 Logical Atomism

8.2.1 Russell: Knowledge by Acquaintance and Description

Analytic Tradition

In many respects logical positivism was motivated by the same desires for perfect clarity and absolute certainty that have driven much of the progress of epistemology from the time of Pythagoras. The particular form that this quest has taken in the *analytic tradition*, which has dominated Anglo-American philosophy in the twentieth century and which includes logical positivism, is the belief that analysis can be pursued down to certain indivisible elements, which are in some way incorrigible, and hence justify the intellectual edifice constructed from them.

Logical Atomism

Knowledge by Acquaintance

One form of this is the theory of *logical atomism* developed by Bertrand Russell in the first decades of the twentieth century.[2] He distinguished *knowledge by acquaintance*, which is indubitable, from *knowledge by description*, which may be in error. Examples of knowledge by acquaintance are sense

[2] The logical atomism of Russell differs from that of Leibniz (p. 117) in that the former is empirical whereas the latter is rationalistic, though, interestingly, both took the *self* to be an elementary given. Sources for this section include Russell (LA) and Russell (*PLA*).

data and the contents of our own memory. So, for example, if I were perceiving a patch of yellow color in a certain place in my visual field, Russell would take that fact to be unassailable, though I could certainly be incorrect in assuming that I'm seeing a yellow object, since it could be a hallucination, I might have jaundice, or be wearing yellow sunglasses, etc. etc. Similarly, my memory that a certain person attended my fifth birthday party might be mistaken, but I cannot be mistaken in the fact of having that recollection.

Knowledge by Description

A common logical fallacy is the use of descriptions to which nothing answers. The classic example is "the present King of France," since France does not have a king. Therefore, when I make an assertion, such as "the present King of France is bald," it is neither true nor false, since it is not an assertion about anything. Such knowledge by description is fallible, since it contains descriptions that may fail in their reference, that is, they do not refer to anything.[3] In contrast, an object of acquaintance acts like a *logically proper name* and cannot fail in its reference; by virtue of my perceiving it, the spot of yellow I'm perceiving refers to itself. For example, I may be incorrect in thinking that I am seeing a dog (I may be hallucinating etc. etc.), but I cannot be incorrect in my belief that I'm seeing a certain spatial arrangement of patches of color and texture ("canoid color patches," as Russell calls them), which I have come to associate with dogs.

The way to certainty, then, lies in the grounding of all descriptions in objects of acquaintance. In the same way that he defined numbers in terms of more basic logical concepts, the usual objects of science would be defined in terms of sense data.[4] For example a "canoid perception" might be defined as an extremely complex logical construction of elementary propositions referring to canoid color patches. In this way we might be absolutely certain about whether we are having a "canoid perception" (though we might be

[3]The pitfalls of reference are illustrated by the footnoted sentence, for had I said "they refer to nothing," it might have implied that there is a thing — Nothing — to which descriptions refer when nothing else answers the description. We could then make a career out of investigating the properties of this Nothing. Indeed, Alexis Meinong (1853–1921) did precisely that, since he gave a kind of logical reality to any *intentional object*, that is, any object of consciousness. Consideration of Meinong's work led Russell to his theory of descriptions.

[4]Or whatever the elementary terms might be; Russell changed his mind more the once in his long career. The nature of the "atomic facts" was also a perennial problem for the logical positivists. I may add in passing that the whole notion, that sense data — such as canoid color patches — are perceptually simple, is open to criticism, and is in fact rejected by the theories advocated in the second half of this book.

mistaken about seeing a dog). Russell was convinced that every meaningful proposition could be analyzed into a description involving objects of acquaintance; in his words:

> Thus in every proposition that we can apprehend (i.e. not only in those whose truth or falsehood we can judge of, but in all that we can think about), all the constituents are really entities with which we have immediate acquaintance. (Russell, OD, p. 55)

8.2.2 Wittgenstein: The Tractatus

> *The limits of my language* mean the limits of my world.
>
> — Wittgenstein (*TLP*, ¶5.6)

> We feel that even when *all possible* scientific questions have been answered, the problems of life remain completely untouched. Of course there are then no questions left, and this itself is the answer.
>
> — Wittgenstein (*TLP*, ¶6.52)

> *Wovon man nicht sprechen kann, darüber muss man schweigen.*
> (Whereof one cannot speak, thereon one must remain silent.)
>
> — Wittgenstein (*TLP*, ¶7)

Ludwig Wittgenstein: 1889–1951

It is rare enough when someone revolutionizes philosophy in their own lifetime; Ludwig Wittgenstein has the distinction of having accomplished it twice. Indeed his views changed so completely that it is common to treat "early Wittgenstein" and "late Wittgenstein" almost like two different philosophers. In this chapter the topic is early Wittgenstein, late Wittgenstein will be considered in Section 11.1.2.

Wittgenstein's career was anything but ordinary. Educated as an engineer, he became interested in mathematics and logic and studied under Russell at Cambridge. He retreated to a primitive home in Norway to think about philosophy, and during the First World War he completed the *Tractatus*, which is the subject of this section. Then he retired from philosophy, since he thought he had solved its principal problems, and held various jobs, including village school-master and gardener in a monastery. However, he began to question his conclusions and eventually rejected most of them. Therefore in 1929 he returned to Cambridge University and philosophy. However, "He believed that being a professor jeopardized the intellectual integrity of a philosopher" (Hartnack, *W&MP*, p. 7), so in 1947 he retired to a cabin on the west coast of Ireland. He returned to Cambridge in 1949 when he was diagnosed with cancer, and died in 1951. Although he published only the *Tractatus*, after his death his students edited and published his manuscripts and the notes he had dictated to his classes.[5]

The Tractatus

Wittgenstein's only published work, the *Tractatus Logico-Philosophicus* (Wittgenstein, *TLP*), had great influence on Russell after World War I, and was seminal to the logical positivist movement. It's style is terse and dogmatic; for example, it begins (Wittgenstein, *TLP*, p. 7):

1 The world is all that is the case.

1.1 The world is the totality of facts, not of things.

1.11 The world is determined by the facts, and by their being *all* the facts.

1.12 For the totality of facts determines what is the case, and also whatever is not the case.

1.13 The facts in logical space are the world.

1.2 The world divides into facts.

Wittgenstein's own one-sentence summary of the *Tractatus* is the oft-quoted:

> What can be said at all can be said clearly, and what we cannot talk about we must pass over in silence.
>
> *Was sich überhaupt sagen lässt, lässt sich klar sagen; und wovon man nicht reden kann, darüber muss man schweigen.*
> (Wittgenstein, *TLP*, p. 3)

Wittgenstein believed that many philosophical problems are a result of misunderstanding language and the way it works (an opinion he did not reject

[5]Hartnack (*W&MP*) is a good, brief introduction to Wittgenstein's thought (both early and late).

later). Language is only capable of expressing certain ideas, and for these we should use a language ideally suited to accuracy of expression. Those things about which we cannot speak, which Wittgenstein called "the mystical" (*das Mystische*), must be passed over in silence, because it is a philosophical error to try to express the inexpressible. He classed ethics and aesthetics in "the mystical." However, Wittgenstein's insistence on the existence of important, inexpressible problems was ignored by the logical positivists.

Atomistic and Picturing Theses

The greater part of the *Tractatus* is devoted to the expressible and its clear expression, which is based on two important theses. The *atomistic thesis* says that analysis terminates in elementary facts; though Wittgenstein's version of this thesis is not the same as Russell's or the logical positivists' I will pass over the details. The *picturing thesis*, which is based on the work of Frege and Russell, says that language, when correctly analyzed, is seen to be "isomorphic" to the world.

2.1 We picture facts to ourselves.

2.11 A picture presents a situation in logical space, the existence and non-existence of states of affairs.

2.12 A picture is a model of reality.
...

2.16 If a fact is to be a picture, it must have something in common with what it depicts.

Russell, in his introduction to the *Tractatus*, explains the picturing thesis as follows:

> In order that a certain sentence should assert a certain fact there must, however the language may be constructed, be something in common between the structure of the sentence and the structure of the fact. (Wittgenstein, *TLP*, p. x)

Thus the structure of the world may be discovered by analyzing the possible logical structure of language that depicts, that is, language that does not attempt to express the inexpressible.

8.3 The Vienna Circle and Verifiability

8.3.1 Background

Moritz Schlick: 1892–1936

Logical positivism was born in the 1920s with the formation of the *Vienna Circle*, a group of physicists, mathematicians and social scientists, who met every week in Vienna under Moritz Schlick, a philosopher at the University of Vienna.[6] The philosophical movement that they started was later known as *logical positivism*, *logical empiricism* and *scientific empiricism*. In part because of the flight of philosophers and scientists from Nazi Germany, logical positivism became the dominant philosophy of science in the Anglo-American world, often more implicitly than explicitly.

Themes of Logical Positivism

Although there were disagreements from the beginning in logical positivism, and they multiplied as the movement progressed, there are several themes that are characteristic. First, it is *empirical* in its requirement that all knowledge (except for the "analytic" truths of mathematics and logic, cf. p. 124) be derived from experience. It is *positivist* in its rejection of "metaphysical" claims that cannot be verified by appeal to the givens (literally, the *data*) of experience. It's *scientific* in holding that the methods of the sciences — especially physics — are the only way to true knowledge. Finally, it's *logical* in recognizing the important role that logical constructions play in relating basic data (sensations, measurements, etc.) to the higher order objects of scientific theories (atoms, waves, energy, etc.). In this way the positivists tried to augment Mach's empiricism with the more mathematical approach of Poincaré. We've seen that Russell and Wittgenstein advocated a similar viewpoint, so it's hardly surprising that the *Tractatus* was treated almost as sacred scripture, and was read out loud at Vienna Circle meetings.[7]

8.3.2 Verifiability and Meaning

Verifiability Principle

One of the central tenets of logical positivism was the *verifiability principle*, which says that the meaning of a proposition lies in its mode of verification. In other words, its meaning consists in the measurements we would have to make to decide whether it is true or false. As Schlick (P&R, p. 87) said:

[6] Among the members of the Vienna Circle were Rudolf Carnap, Otto Neurath, Friedrich Waismann, Kurt Gödel, Phillip Frank, Karl Menger, Hans Hahn and A. J. Ayer.

[7] For a time the prophet attended the meetings in person, but in the end he was less welcome than his revelations.

> The criterion of the truth or falsity of the proposition then lies in the fact that under definite conditions (given in the definition) certain data are present, or not present. If this is determined then everything asserted by the proposition is determined, and I know its meaning.

Example: Verification Conditions of Temperature

For example, consider the proposition that the temperature of a certain volume of water is 90°C. What does it mean? The logical positivist answer is that if you put a thermometer into the water and read it, then it will show 90°C. However, "reading a thermometer" is still a rather vague description, so it is more precise to say that we place a Celsius thermometer in the water, wait for its reading to stabilize, and then observe that it's opposite the '90' mark. This is still somewhat far from sense data, however, so a further explication might be: "When the meniscus of the mercury column stops moving, it is closer to line marked '90' than to the lines marked '89' or '91'." In general, the idea is that all meaningful factual statements can be reduced to measurements, and measurements are ultimately simple perceptual judgements, such as determining which mark on a dial is the closest to a pointer.

Rejection of "Metaphysics"

A corollary of this reduction of meaning to verification conditions is the conclusion that any proposition that cannot be verified by appeal to data is *literally meaningless*. In Schlick's words (P&R, p. 88),

> A proposition which is such that the world remains the same whether it be true or false simply says nothing about the world; it is empty and communicates nothing; I can give it no meaning.

Since traditionally *metaphysics* is the study of the basic grounds of existence, which is taken to be prior to empirical verification, its fails the verifiability test, and among the positivists "metaphysics" became a term of derision for meaningless mumbo-jumbo. In this way they saw the verifiability principle as a razor that would cut away all the unanswerable — because meaningless — questions that had vexed philosophers for ages, such as the existence of the external world. It would not *solve* metaphysical problems, it would *dissolve* them.

Existence of Atoms

For a specific example, in the early days of atomic theory, when the existence of atoms was inferred from chemical laws, there were vigorous debates about the reality of atoms. Were there really indivisible bits of matter, or were they simply a convenient fiction for calculation?[8] Though many of the senior scientists at the time were against the reality of atoms, the tide eventually turned against them, and modern imaging techniques make it nearly impossible to doubt their existence. The positivist answer would be: "Tell me what you mean by 'existence' by reducing it to measurement. If you do so, then we can make the measurements and answer your question. If you can't point to any observable differences resulting from the existence or nonexistence of atoms, then you are just making meaningless sounds. The solution to your problem is for you to see the emptiness of your question."

Positivist View of Ethics and Aesthetics

Some logical positivists thought that ethical and aesthetic judgements have no more meaning than emotional ejaculations. Thus an aesthetic judgement such as "That's beautiful" has the same content as "Wow!" or "Nice!" An ethical judgement, such as "That's wrong," is really just an expression of personal feeling, like a snarl of anger, a hurt cry, or a sympathetic moan. If these have any meaning at all, it is as reflections of the speaker's emotional state, which is to be reduced to physiological measurements. Thus A. J. Ayer (1910–1989) says:

> For we have seen that, as ethical judgements are mere expressions of feeling, there can be no way of determining the validity of any ethical system, and, indeed, no sense in asking whether any such system is true. All that one may legitimately enquire in this connection is, What are the moral habits of a given person or group of people, and what causes them to have precisely those habits and feelings? (Ayer, *LTL*, p. 112)
>
> Such aesthetic words as "beautiful" and "hideous" are employed, as ethical words are employed, not to make statements of fact, but simply to express certain feelings and evoke a certain response. It follows, as in ethics, that there is no sense in attributing objective validity to aesthetic judgements, and no possibility of arguing about questions of value in aesthetics, but only about questions of fact. A scientific treatment of aesthetics would show us what in general were the causes of aesthetic feeling, why various societies produced and admired the works of art they did, why taste varies as it does within a given society, and so forth. (Ayer, *LTL*, p. 113)

[8] In more recent times there have been similar debates about the reality of quarks.

8.3.3 Reductionism and the Unity of Science

Another important component of logical positivism was the *unity of science movement*.[9] Carnap gave a systematic explanation of how, in general, propositions expressed in the terms of one science, say biology, could be reformulated into equivalent propositions in the terms of another, say physics. Although the propositions of physics could conceivably be reduced to those of biology as well as the other way, in practice it was assumed that physics would be the ultimate goal of all reductions:

> The thesis of *physicalism* maintains that the physical language is a universal language of science — that is to say, that every language of any sub-domain of science can be equipollently translated into the physical language. From this it follows that science is unitary system within which there are no fundamentally diverse object-domains, and consequently no gulf, for example, between natural and psychological sciences. This is the thesis of the *unity of science*. (Carnap, *LSL*, p. 320)

Although the logical positivists were never able to carry out this program, their legacy lingers on in many sciences in the form of a desire to explain their phenomena in physical terms, and of an acceptance of the language and methods of physics as the standards of "real science" (an attitude sometimes called "physics envy" — with acknowledged Freudian allusions).

[9] Carnap (LFUS) is a clear explanation of the unity of science movement and the methods of reductionism.

8.4 The Collapse of Logical Positivism

The End of Logical Positivism

Already by the 1930s the logical positivist movement had begun to fragment. Although this was the beginning of the end, the movement continued for several decades, and its attitudes and methods linger to this day in many disciplines. The causes of the collapse of logical positivism can be put in two classes, *internal weakness* and *external criticism.* I will begin with the internal causes.

8.4.1 Nature of Atomic Facts

The Problem of Atomic Facts

There were problems in the logical positivist program from the beginning, as there are in all new paradigms, but they were expected to be easy to solve. As it turned out, they were fatal flaws. The first of these problems was the exact nature of the "atomic facts," which were supposed to be the indubitable raw material of scientific theories. I've mentioned that Russell changed his mind on this issue; the logical positivists were also unable to reach a consensus.

Schlick: Relations Among Sense Data

At first Schlick took sense data to be the atoms, but he recognized that sensory experiences are private — there is no way to tell if the yellow I see is the same yellow that you see — so subjective sense data seemed a bad foundation on which to build a supposedly public, objective science. Instead he decided to start from *relations* among sensory experiences; though our subjective experiences of colors might differ, at least we might be expected to agree about whether two patches have the same color or not. In this way the *order relations* among (private) sense data became the raw material of theories; as Schlick observed, you might even perceive as pitches what I perceive as colors, but that wouldn't matter so long as the relations among your perceptions corresponded to those among mine.

Neurath: Protocol Sentences

However, other members of the Vienna Circle were not comfortable with the abstractness of Schlick's solution, since they wanted empirical science to be based on concrete observations. Otto Neurath (1882–1945) proposed *protocol sentences* as the atomic constituents of knowledge. These were intended as incorrigible statements of facts which would form the basis of theory construction. This is one of his examples:

> Otto's protocol at 3:17 o'clock: [At 3:16 o'clock Otto said to himself: (at 3:15 o'clock there was a table in the room perceived by Otto)]. (Neurath, PS, p. 202)

Protocol sentences such as this seem too complex to be the "atomic constituents of knowledge." Nevertheless, Neurath admitted that protocol sentences are far from unambiguous. He went so far as to suggest that a term such as "Otto" could be defined as the man "whose carefully taken photograph is listed no. 16 in the file"! Of course, though this may reduce the ambiguity, it doesn't eliminate it, and in the end Neurath had to admit that his protocol sentences are not incorrigible. He thought the truth of a protocol sentence could be judged only by its coherence with other protocol sentences and with the theories constructed from them. Such a *coherence theory of truth* was anathema to the Vienna Circle.[10] Therefore protocol sentences were judged insufficient for the rock-solid foundation desired by the positivists.

Schlick: Confirmation Statements

I'll mention just one more stage in this development, to impress on you the elusiveness of "atomic facts." Schlick countered Neurath's proposal with one of his own. We have seen an incompatibility between incorrigibility and public verifiability. On one hand, we may accept that personal sensory experience is indubitable, but it's absolutely private. On the other hand, protocol statements are publicly verifiable, but are inherently ambiguous and potentially in error. Schlick (FK) tried to combine the two by saying that a protocol sentence, such as

> M.S. perceived blue on the *n*th of April 1934 at such and such a time and such and such a place,

is just a *hypothesis*, that cannot be counted as knowledge unless it has a corresponding *confirmation sentence*, such as "Here now blue." However, as Schlick himself observed, a confirmation statement cannot be written down, since "token-reflexive" words such as "here" and "now" lose there meaning. But filling in a description of the time and place converts it into a protocol sentence and incorrigibility is lost. Thus Schlick's analysis permits the individual scientist to have his protocol sentences confirmed (by *private* affirmations such as "Here now blue"), but this is not a satisfactory basis for *publicly* verifiable science.

[10] In roughest terms, a coherence theory of truth says that a system of laws, facts, etc. are true if they hang together well. Crudely, if it's a good story then it's true. Scientists as a rule are *realists*, i.e., they believe there is a real world independent of us as observers, and therefore a statement should be judged true only if it *corresponds* to reality. Crudely, in a *correspondence theory of truth*, a true sentence says what's so.

8.4.2 Verifiability Problems

Analytic and Synthetic Statements

Logical positivism's second major internal problem was the status of the verifiability principle. The positivists held that there were only two classes of meaningful statements. On the one hand were the analytic statements, which express the truths of logic and mathematics. Since they follow necessarily from the axioms and definitions, they are indubitable, but factually empty, for they are consequences of the axioms and definitions, which may be chosen at will. On the other hand are synthetic statements, whose truth depends on their verification. In rough terms we have the truths of mathematics and the truths of science. Anything that could not be put into either of these categories was necessarily metaphysical mumbo-jumbo.

Status of the Verifiability Principle

The trouble with the positivists' two-way classification of meaningful statements was that it seemed to leave no place for the verifiability principle itself! On one hand, it didn't seem to be analytic, and in any case it wouldn't be very convincing as a prescription if it were merely a consequence of arbitrarily chosen axioms and definitions. On the other hand, it could be considered synthetic only if its meaning could be explicated in terms of verification conditions. What sort of measurements or observations could confirm the verifiability principle. What if the observations, when made, *disconfirmed* the principle? The positivists were also unhappy with this possibility.

One solution would be to conclude that the verifiability principle is a meaningful statement that is neither analytic nor synthetic. But to admit one meaningful statement that's neither analytic nor synthetic is tantamount to admitting the meaningfulness of all the banished claims of metaphysics. The only alternative was to admit that the principle is factually meaningless and simply "expresses certain feelings and evokes a certain response."

Carnap Material and Formal Modes of Speech

For example, Rudolf Carnap (1891–1970), one of the major figures of logical positivism (see Carnap, *LSL*, §§64, 74–79), distinguished statements that are in the *material mode of speech*, that is, about the world, from those in the *formal mode of speech*, that is, about language; the distinction is essentially the same as that between *object language* and *metalanguage* (cf. medieval ideas of *intention* and *supposition*, p. 71). The resolution of the problem of the status of the verifiability principle then lay in recognizing that it's a statement in the formal mode of speech, specifically, it is a *recommendation* about how to use the words 'meaningful' and 'meaningless'. If you accept this recommendation, then you are bound by the verifiability principle and the rest of the positivist view. On the other hand, if you do not accept it, then you have simply adopted another idea of meaningfulness — you have chosen to use 'meaningful' and 'meaningless' in a different way from the positivists. But this is no excuse for them to assert that you are speaking nonsense. Carnap expressed this in his *Principle of Tolerance*: "It is not our business to set up prohibitions, but to arrive at conventions" (Carnap, *LSL*, p. 51). In the end the aggressive "If you can't measure it then you don't know what you're talking about" had been replaced by the wishy-washy "We recommend that you always base your definitions on observables."

Are Scientific Laws Verifiable?

Another problem with the verifiability principle is that its strict application would imply that *no scientific law is meaningful!* For example, we may verify that Angus is mortal, that Bridget is mortal and that Camille is mortal, but verification of 'all people are mortal' is impossible, since it would require observation of an infinite number of cases. The major value of scientific laws lies in their applicability to an infinite number of situations, yet it is precisely that property that makes them strictly unverifiable. The logical positivists tried to escape this predicament in many ways, but none were satisfactory. For example some (e.g. Schlick at one point) said that in fact scientific laws are technically *nonsense*, though "important nonsense." This was not a popular position, however, and other positivists distinguished confirmation (or weak verification) from (strong) verification, the idea being that the atomic facts can be strongly verified, and so are incorrigible; whereas general laws can be confirmed by supporting observations, but are not beyond doubt and are always subject to empirical refutation. Unfortunately this position required giving up the certainty that positivism had promised. Indeed Ayer (*LTL*, p. 38) went so far as to declare in 1936 that "no proposition, other than a tautology, can be anything more than a probable hypothesis." Logical positivism had come a long way![11]

What Does 'Verifiable' Mean?

Another problem with the verifiability principle was its exact interpretation. 'Verifiable' means 'possible to verify', but what is the appropriate sense of possibility? Logical possibility? Physical possibility? Economic possibility? For example, Schlick raised the question of the verifiability of the proposition that there are mountains on the far side of the moon. Although at that time there were no rockets capable of circumnavigating the moon, most positivists agreed that the proposition was verifiable, since they could see that it was verifiable *in principle*. That is, the necessary observations are "theoretically conceivable," but the question remains of how far

[11] See Ayer (*LTL*, pp. 36–39) for further discussion of these issues.

this theoretical conceivability can go. For example, I can conceive of devices that use no energy (such as "Maxwell's demon"), and therefore are physically impossible. Can my verification conditions make use of "theoretically conceivable" procedures that violate the laws of physics?

If our verification procedures are restricted to physically possible, then we have the peculiar situation that the meaningfulness of a sentence depends on the state of physical knowledge. For example, before 1927 (when Heisenberg's Uncertainty Principle was published) the proposition that a particular particle has a certain specific position and momentum would have been meaningful, since in principle position and momentum could be measured to arbitrary accuracy. However, in 1927 the sentence became meaningless, since the Uncertainty Principle says that the supposed verification procedure is physically impossible.[12]

Whenever one sees the suffix '-able' (or '-ability') it's worthwhile to ask, "What is the relevant sense of possibility?"

8.4.3 External Criticism

Sympathetic Doubt Concerning Atomic Facts

One must acknowledge the intellectual integrity and honesty of the logical positivists and their allies, for it suffered as much from their criticism as from that of its enemies. We've seen that from the beginning there was disagreement about what should be accepted as the incorrigible givens — the *data* — that were intended to be the atomic facts from which true knowledge would be constructed. In addition to the continuing investigations of Russell, Carnap and others, the idea of atomic facts was scrutinized in the 1930s and '40s by such sympathetic philosophers as Tarski, W. V. O. Quine (1908–), Nelson Goodman (1906–) and Hilary Putnam (1926–). However, we pass over these criticisms, since they are subsumed by the more fundamental critique originating in the phenomenological tradition, which will be considered in Chapter 11.3.

Ordinary Language Philosophy

An even more damaging assault came from the "late Wittgenstein," J. L. Austin (1911–1960), and other "ordinary language philosophers." Since

[12] One also wonders about the precise way in which the meaning vanishes. Does it disappear as soon as the Uncertainty Principle is formulated, or only when it becomes widely accepted? Or does it vanish for each physicist as they accept the principle (which would make 'meaningful' subjective)? Or do we discover that the proposition was *never* meaningful (which would mean we can rarely be sure of meaningfulness, since we may revise at any time our beliefs about physical possibility).

they will be taken up in detail later (Ch. 11.1.2), it will suffice here to say that they called into question the view of language that had been accepted through most of the Western philosophical tradition. First, they pointed out that the words of ordinary language do not have definitions in terms of essential characteristics, as was generally assumed, and further that their meaning could not be captured by such definitions. Therefore, the method of logical construction was fundamentally inadequate as a means of explicating "metaphysical" questions relating to the mind, consciousness, perception, and so forth. Second, they showed that the meaning of a statement in ordinary language is not a simple function of the things denoted by its elementary terms (as, for example, in Tarski's semantic theory, Sec. 6.2). Although each statement fulfills a function for the community that uses it, that function might not be the simple "picturing" of a state of affairs (as "early Wittgenstein" had held). Although these conclusions followed from an analysis of ordinary language, it was apparent that they applied equally well to much of scientific language (see Ch. 11.4.4), so whatever logical positivism might be doing, it was *not* explaining the *actual* practice of science or extending it to other disciplines. Finally, the deepened understanding of *meaning* that came from ordinary language philosophy showed that the much-maligned "metaphysical" statements are far from meaningless. As Putnam (*MLR*, Vol. 2, p. 20) sadly remarked,

> Not a single one of the great positive theses of Logical Empiricism (that Meaning is Method of Verification; that metaphysical propositions are literally without sense; that Mathematics is True by Convention) has turned out to be correct.

8.4.4 Summary

So we may say that logical positivism came in with the bang of the *Tractatus*, but went out with the whimper of increasingly rarefied debates about atomic facts and the status of the verifiability principle. Perhaps because its quick birth was more visible than its slow death, remnants of logical positivism have lingered on in the sciences long after most philosophers abandoned it.

8.5 Influence of Logical Positivism

8.5.1 Quantum Mechanics

> The Pythagorean school was an offshoot of Orphism, which goes back to the worship of Dionysus. Here has been established the connection between religion and mathematics which ever since has exerted the strongest influence on human thought. The Pythagoreans seem to have been the first to realize the creative force inherent in mathematical formulations.
>
> — Heisenberg (*P&P*, p. 67–68)

Heisenberg's Uncertainty Principle

The influence of logical positivist ideas is apparent in the *Copenhagen Interpretation* of quantum mechanics, named for Bohr's Copenhagen Institute of Theoretical Physics, where it was developed in the mid-1920s, mostly by Niels Bohr (1885–1962) and Werner Heisenberg (1901–1976). The Heisenberg Uncertainty Principle, one of the pillars of quantum mechanics, says that there is an unavoidable uncertainty in the measurement of certain "conjugate variables," such as position and momentum, or time and energy.[13] Specifically, if Δx is the uncertainty in the position of a particle and Δp is the uncertainty in its momentum, then[14]

$$\Delta x \, \Delta p \geq h.$$

Therefore, if we increase the accuracy of our momentum measurement, we must "pay for it" with decreased accuracy in the position measurement and vice versa.[15]

The Copenhagen Interpretation

[13] There is of course an enormous literature on the Uncertainty Principle, both popular and technical, by both physicists and philosophers. My goal here is limited to illustrating the ubiquity of logical positivist thinking in science. Although quantum mechanics is not within the scope of this book, it's perhaps worth remarking in passing that the uncertainty relation can be traced to the desire to have consistent discrete (particle) and continuous (wave) descriptions of physical phenomena. See Heisenberg (*P&P*, Ch. 3) for a readable discussion of the Copenhagen Interpretation.

[14] The exact constant, Planck's constant h here, depends on the way the measurement process is defined.

[15] It's often thought that this uncertainty results from the measuring device "disturbing the system," but this is not the case, for the uncertainty relation has been shown to hold

Although the Uncertainty Principle has been very well confirmed since it was first proposed in 1927, physicists still disagree about how it is to be understood. The common-sense view, which we might call "naive realist," is that a particle actually has some specific position and momentum, but that nature places limits on their joint measurement. This, however, is not the view of most physicists, who endorse instead the Copenhagen Interpretation, which says that since it is *in principle* impossible to measure an exact position and momentum, the particle does not in fact have them. This can be understood as a direct application of the Verifiability Principle. According to the laws of physics (i.e. Heisenberg's principle), it is impossible to verify a particle's exact position and momentum, so it is literally *meaningless* to speak of it having an exact position and momentum. We can speak meaningfully only about what we can measure. Since we can measure only an approximate position and momentum, it only makes sense to speak of an approximate position and momentum (as determined by the laws of quantum mechanics).

By taking positivism to its logical — if absurd — conclusion, respected physicists have gone so far as to claim that it's meaningless to say that the moon is still there when you are not looking at it. (How could you observe that the moon is there when it's not being observed?)[16]

---

even when the measurement is made by the *absence* of an interaction (so-called Renninger-style measurements). In fact, the uncertainty relationship is an inevitable consequence of the mathematics of quantum mechanics, since it is a theorem of Fourier analysis that any Fourier-transform pair of variables (e.g. time and frequency) will satisfy such a relationship (Gabor, TC).

[16]The difficulty physicists have accounting for the "collapse of the wavefunction" is also a consequence of their positivist approach with its dualist perspective. For logical positivism separates the subject from the object, the observer from what is observed. As Heisenberg (*P&P*, p. 55) states explicitly, quantum mechanics "starts from the division of the world into the 'object' and the rest of the world, and from the fact that at least for the rest of the world we use the classical concepts in our description."

Since by the Copenhagen Interpretation a particle cannot be said to have a definite state until its *observed* to have a definite state, it is said that before the observation the particle exists as a "superposition" of possible states, but that after the observation this superposition collapses into a specific state (corresponding to what was actually observed). It must be emphasized that this collapse is required only to account for the definiteness of observations. The standard dogma is that so long as a physical system is unobserved, it continues to evolve as a superposition of possible histories. On the other hand, since the laws of physics are supposed to apply to everything in the universe, including observers and their observations, one might expect there to be a physical process corresponding to

Positivism in Relativity Theory

Although Albert Einstein (1879–1955) disagreed with logical positivism, and he published his first papers on special and general relativity (1905, 1916) before the formation of the Vienna Circle, nevertheless we can see in them applications of the same idea, that physical concepts are constructions of observations and that unverifiable differences are not real differences. In 1916 Einstein directly acknowledged his debt to Mach, saying "the study of Mach has been directly and indirectly a great help in my work," and "Mach recognized the weak spots in classical mechanics and was not very far from requiring a general theory of relativity half a century ago" (quoted in Newman, *WM*, Vol. 3, p. 1785). For example, the postulate of relativity states there is no experiment that we can perform to distinguish one reference frame from another that is moving at a constant velocity with respect to the first. That is, the terms 'absolutely at rest' and 'absolutely in motion' are unverifiable and hence meaningless; all that can be verified is relative motion. Similarly, general relativity is based on the postulate that an accelerating reference frame is experimentally indistinguishable from a reference frame in a gravitational field. This opens the way for a geometrical description of gravity in terms of the curvature of space.

8.5.2 Behaviorism

> Behaviour is simply part of the biology of the organism and can be integrated with the rest of the field when described in appropriate

the collapse of a superposition of potential observations into a single actual observation.

This view generates a host of problems. What is the nature of this collapse? If it is a physical process, what causes it? If it's caused by the act of observation, what constitutes an observation? Must the observation be made by a human? Or might an animal's awareness be adequate to cause the collapse? Could a machine make the observation? This line of reasoning has led to incredible conclusions, such as that consciousness had to evolve in the universe so that it would have a definite state! It also leads to a number of paradoxes, of which there are many popular accounts. These paradoxes vanish once one completely discards the dualism implicit in a fundamental distinction between the observer and the observed (MacLennan, DD).

physical dimensions.

— Skinner (BSO, p. 75)

Connection with Logical Positivism

Further evidence of the influence of logical positivist thinking is apparent in *behaviorism*, the philosophy of psychology that dominated that field through the first half of the twentieth century, and still exerts a strong influence. Although the first "behaviorist manifesto" (Watson, 1913) predates the Vienna Circle, B. F. Skinner (1904–1992?), one of the principal exponents of behaviorism, traces its ideas to Mach, Poincaré and Comte — the same progenitors claimed by the logical positivists (Skinner, BSO, p. 74). Certainly behaviorism and logical positivism both reflect ideas "in the air" at the time.

John B. Watson: 1878–1958

The founder of behaviorism was John B. Watson, who became dissatisfied with psychology's "introspective" methods while he was a doctoral student in the first decade of the twentieth century. At that time psychology, as an experimental science distinct from philosophy, had existed for only 25 years, and its principal method was to ask observers to report on their mental states, which they had "observed introspectively." Watson thought that psychology could achieve its goals better if it studied people the same way it studied animals: by observing overt behavior. In this way psychology would be more scientific, since it would make publicly testable predictions of behavior, and it would be more practical, since it would produce means for controlling behavior. Watson described his vision for a new psychology in summer lectures at Columbia in 1912; they were published the following year (Watson, 1913). Although there were strong objections from some quarters, behaviorism advanced quickly. Its hegemony over psychology from the 1920s to the 1950s can be attributed in part to its conformity with the contemporary positivist philosophy of science, but also to it's arrival on the scene just in time to stock the rapidly growing psychology departments of the '20s and '30s.[17]

Rejection of Mental Phenomena

[17] Watson was left out of much of behaviorism's success. As a result of a divorce scandal, he was shunned by academia and forced to earn a living selling rubber boots. Eventually he was hired by an advertising firm, where he applied his knowledge of behavior to the development of many of the techniques of modern advertising. Behaviorism's advocacy of and technology for behavior control has inspired at least three novels: Huxley's *Brave New World*, Orwell's *1984* and Skinner's *Walden Two*. (See *OCM*, s.v. Behaviorism, and Cohen, *JBW*.)

Behaviorism's emphasis on publicly observable behavior and its rejection of introspection of mental states has much in common with logical positivism's rejection of "metaphysics" in favor of measurement. For example, Skinner (BSO, p. 75) says,

> By dismissing mental states and processes as metaphors or fictions, [behaviorism] directs attention to the genetic and personal histories of the individual and to the current environment, where the real causes of behavior are to be found.

Mental experiences, such as a feeling of purpose, are *epiphenomena* with no causal efficacy, and the cause of our actions is to be found in the environment, past or present:

> We do not act because we have a purpose, we act because of past consequences which generate the condition of our bodies which we observe introspectively and call felt purpose. (Skinner, BSO, p. 74)

Measuring the Mind

"Physics envy" brings with it the assumption that everything of importance can be measured, so it is no surprise that Edward Lee Thorndike (1874–1949), who published his *Measurement of Intelligence* in 1926, was a leading behaviorist. Of course, there is nothing wrong with trying to measure phenomena. The problems arise when measurement is combined with the positivist outlook, which tends to reject the unmeasurable (or even the unmeasured) as meaningless and unimportant. Then, the measured fragments of the phenomenon come to be viewed as the whole of the phenomenon. For example, creativity seems to have little correlation with the components of intelligence measured by IQ tests (*OCM*, s.v. Creativity), but IQ is frequently treated as an accurate measure of intelligence.[18]

Reduction of Behavior to Rules

The positivist's emphasis on logical construction, for example in Carnap's "logical syntax," is also manifest in behaviorist psychology. For example, Clark Leonard Hull (1884-1952), the leading behaviorist of the 1940s,

> wanted a theory of behaviour as formal as Euclid's theory of geometry, and he introduced theorems and postulates which could account for all behaviour. (*OCM*, s.v. Behaviorism, p. 73)

[18] See Gould (*MM*, esp. Chs. 5, 6) for a discussion of the abuse of intelligence measurements.

(His intended "behavior theory" was not successful.) Debate continues about whether behaviorism improved psychology or merely delayed its progress. Be that as it may, in the next chapter we will consider *cognitive science*, a more sophisticated application of atomism and logical construction in psychology.

8.5.3 The Turing Test

> I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning.
>
> — A. M. Turing (CM&I, 1950)

Debate continues about whether a computer could think or have a mind, and conversely about whether the brain or mind is in some sense a computer. Often the discussion revolves around the *Turing Test*, which Turing called the *imitation game* (Turing, CM&I). The topic here is limited to the Turing Test as a reflection of logical positivism; more general considerations will be taken up later (**??**).

Description of the Turing Test

Can machines think? This is the question with which Turing begins, but he observes that the terms are not clearly defined. He entertains the possibility of capturing the everyday use of these terms in definitions — essentially the method of ordinary language philosophy — but concludes that the result would have no more significance than a public opinion poll. Turing concludes:

> But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous terms.

The new question is based on the imitation game: First consider a game played by a man, a woman, and an interrogator of either sex. The "contestants" are concealed from the interrogator, who can communicate with them only by typed messages. The object of the game for the interrogator to guess which contestant is the woman; the man's object is to fool the interrogator into thinking he is the woman; the woman's object is to convince the interrogator that she is in fact the woman. We can expect the interrogator to reach the correct conclusion a certain fraction of the time (probably more than half). Now, Turing says, suppose we replace the male contestant by a computer, so that both the computer and the woman are trying to convince the interrogator that they are the woman. Turing says the original question, "Can machines think?" can be replaced by the question, "Can a machine fool the interrogator at least as often as a human male contestant?"[19]

Evaluation of the Turing Test

As evident in the quotation that begins this section, Turing thought that by the beginning of the Twenty-first Century a computer would pass the Turing test, but we disregard that for now. Here our concern is not the *answer* to the question, but how logical positivism influenced the *formulation* of the question.

First observe that the Turing Test is completely behavioristic, replacing the mental predicate "think" by a verifiable behavioral predicate, "able to win the imitation game at least as often as a human male."

Turing acknowledges that passing the Turing test does not prove that the machine is conscious, but he counters this with the well-known claim that we do not know if other *people* are conscious. He observes that "Instead of arguing continually over this point it is usual to have the polite convention that everyone thinks." He anticipates that, rather than retreating to a solipsistic position, most people would be willing to extend this polite convention to a Turing Test-passing machine.

In other words, Turing takes consciousness to be a property that is not publicly verifiable and that is accessible only through introspection. Thus it has no role in his verifiable surrogate for "Can machines think?"[20]

[19] Popular accounts of the Turing Test often have a person and a machine both trying to convince the interrogator that they are the person; the sexual component is omitted. Though Turing does not explain the purpose for this extra complication, and most descriptions of the Turing test simply ignore it, presumably it is to make the test more fair by comparing the intelligence of two agents in comparable tasks — imitating what they are not. That is, the goal is for the machine to know as much about being a woman as a man knows about being a woman, but not necessarily as much as a woman knows about being a woman.

[20] Turing's original paper (Turing, CM&I) addresses many (though not all) of the objections that have been made against the Turing Test and machine intelligence over the years, and anyone interested in the topic would be well advised to read it, and not depend on this or any other second-hand account.

8.5.4 Conclusions

Logical Positivism had Many Attractive Features

The basic goals of logical positivism are laudable. "What we can say we should say clearly" — who would argue with this?[21] Likewise, a question can often be clarified by asking: How would you know? Or: What difference would it make? And shouldn't we try to relate our knowledge in one discipline to that in another, and ferret out any inconsistencies? These are common values in all intellectual activities, but especially in science.

Excesses Resulted from Quest for Certainty

The problems arise when these goals are pursued to extremes — one consequence of the quest for certainty. For the logical positivists wanted was an absolute, context-free foundation for knowledge. In this they were no different from the rationalists of earlier times. As empiricists, they differed in taking the objects of sense as their starting point, rather than the objects of the intellect. They were alike in supposing they knew where to find absolute certainties, but when they tried to sieze them, they slipped through their fingers.

Observation is Always in Context

The logical positivists ignored that fact that observation takes places in a context, a dense web of common practices (e.g. measurement) and unstated assumptions (e.g. causality). As a consequence, there are no context-independent sense data, nor are there observations unbiased by preconceived ideas. Further, most significant concepts are not logically reducible to necessary and sufficient properties equivalent, in turn, to their verification conditions.[22] Science progresses in spite of these difficulties, but they undermined the foundations of logical positivism. Though they ridiculed philosophy's quest for metaphysical truth, the logical positivists fell prey to an equally alluring idea and naively assumed the existence of context-free, atomic facts. Nietzsche foresaw the trap:

> One must, however, go still further, and also declare war, relentless war unto death, against the "atomistic need" which still leads a dangerous afterlife in places where no one suspects it, just like the more celebrated "metaphysical need"... (Nietzsche, *BG&E*, §12)

Negative Consequences of Positivism

[21] In fact there are situations, both social and legal, in which intentional ambiguity is desirable. We ignore this for now, and restrict our attention to scientific and philosophical discourse, where presumably clarity is always desirable.

[22] These issues will be discussed in more detail in Chap. 11.4.4.

One unfortunate consequence of logical positivism, which still infects those disciplines where it has been influential, is the dogmatic, intolerant, arrogant and closed-minded attitude that it encourages. For when people are certain they are correct, they tend to be intolerant of other points of view. Why compromise with falsehood? And so we often find physics envy reinforced by *physics arrogance*: If you don't do things like we physicists (say we) do things, then you aren't a real scientist.[23]

Ignoring the Unquantified

One common form taken by this arrogance is the dogmatic limitation of phenomena to the measurable. Certainly, there is no inherent harm in trying to measure things; the mistake comes in the assumption that such observational or pragmatic correlates exhaust the phenomena. Thus we hear, "Intelligence is what IQ measures." Other important but unmeasured components of intelligence are simply defined out of existence, and therefore disvalued.

In conclusion, the major value of logical positivism, as of rationalism, formalism, and other single-minded views, may be that it exposed its own limitations. Now that we have thoroughly explored *that* path, we can turn our attention to others.

[23]As a consequence, many "soft sciences" and would-be sciences try to emulate the methods of physics to excess. For example, computer science has engendered the (largely unfruitful) "software physics" and "software science."

Chapter 9

The Computational Theory of Mind

9.1 Historical Background

9.1.1 Gestation

Recap: Arithmetization and Formalization

We've seen that for millennia the idea has persisted that comprehensibility ultimately requires reduction to finite arrangements of discrete tokens. In mathematics this notion manifested itself first in the attempt to arithmetize geometry, that is, to reduce the continuum to discrete points. Secondly the preference for the discrete led to the attempt to formalize knowledge, that is, to reduce knowledge to formulas, finite arrangements of tokens, and inference to the mechanical rearrangement of formulas.

Formalization of Knowledge

Traditionally there were taken to be two kinds of knowledge, analytic and synthetic. As defined by the logical positivists, analytic knowledge was exemplified by the arbitrary, factually empty, but useful truths of formal mathematics; synthetic knowledge comprised the empirical truths of "real science." The former took the ultimate source of truth to be the (freely chosen) axioms, the latter took it to be atomic facts (whatever they might be). Both used formal logic to construct more complex ideas from more basic ones.

Formalization of Thought

The idea that knowledge can be expressed as a formal system naturally suggests that thought is the mechanical rearrangement of finite structures of discrete tokens, that is, that thought is computation. We've seen different versions of this idea in the work of Lull, Hobbes, Leibnitz, and Jevons.

We have followed the attempt to formalize knowledge in preceding chapters; in this chapter we pick up the trail of the formalization of thought, which eventuated in a new discipline, *cognitive science*.

Stimulus From World War II

Although, as we have seen, there was no lack of research into the formalization of thought up through the 1930s, the pace was accelerated by the needs of the Second World War. Increasingly sophisticated weapons, including the atomic bomb, required complex computations and "intelligent" control systems. The result was an increase both in resources – human and financial – and in emotional commitment.[1]

Development of Electronic Computers

The most tangible result was the development of the first practical electronic computers (e.g. ENIAC: 1946, EDSAC: 1949, EDVAC: 1950), which were needed for calculating numerical tables. The possibility of manipulating nonnumerical information was demonstrated by Turing's cryptanalysis work, for which he used computers not unlike Turing machines (Sect. 6.4.2). The war effort also led to important theoretical advances, such as the development of *cybernetics* by Norbert Wiener (1894–1964). Cybernetics can be defined as the study of control systems in animals and machines; in Europe the term *cybernetics* includes artificial intelligence, though in the United States it's used more narrowly. Further, in an influential book, *Design for a Brain* (1952), W. Ross Ashby (1903–1972) used formal methods to show how a computational system could learn.

von Neumann: 1903–1957

Throughout the '40s and '50s the influence of von Neumann was pervasive; he was active in computer design, automata theory, formal mathematics, neuroscience and cybernetics. His broad viewpoint helped to keep together the many strands — practical, theoretical, technological, biological, mathematical and philosophical. As much as any individual he represents the coalition of interests that was to constitute cognitive science.

Claude Shannon: 1916–2001

Claude Shannon, who in 1938 had shown the relation between digital circuits and Boolean logic, went on to develop *information theory* with Warren Weaver (1894–1978) in 1949. This theory, which perhaps stemmed from insights gained in wartime cryptographic work, is positivist in method, for it reduces information to the probability of identifying symbols, that is, to a behavioral criterion. It has been very influential in many sciences and even in the humanities, though this may be more an consequence of its suggestive

[1]Much of the historical information in this section comes from Gardner (*MNS*), which is a comprehensive presentation of the history of Cognitive Science. Chapters 2 and 3 are especially relevant to this section.

name rather than a close connection to our everyday idea of information. For example, it has been pointed out that the book (of a given size) containing the most information (in Shannon's sense) is a book of random numbers (Hamming, *C&IT*, p. 103).[2] One again we see the positivist method — take a concept, such as information, which no one knows how to quantify — and replace it by a much simpler quantifiable concept, such as prediction probability; and we see the typical consequence of the method, for many investigators, even in disciplines where they should know better, no longer recognize a difference between the ordinary and Shannon notions of information. In effect, information has contracted to Shannon's definition.

An Important Symposium in 1948

An important event was a 1948 symposium on "Cerebral Mechanisms in Behavior," which brought together many of the new ideas. For example, von Neumann discussed the similarities between the computer and the brain. Warren McCulloch (1898–1969) developed a similar theme, based on prior work with Walter Pitts (1923–1969), and showed how simple, neuron-like computing elements could function like the logic circuits of computers.

Karl Lashley: 1890–1958

Most importantly, however, the neuropsychologist Karl Lashley gave convincing arguments why behaviorism was inadequate. Behaviorist theory was organized around the reflex arc, but Lashley argued that this could not account for complex, organized behavior, such as planning, reasoning, problem solving and especially language use. Thus, Lashley's paper signaled the beginning of a shift of emphasis from "lower" cognitive functions, such as reinforcement of stimulus-response pairs, to "higher" cognitive functions, such as game playing, problem solving, abstract reason and the use of language.[3] Structured behavior of this kind is, seemingly, the manipulation of discrete symbols, so the new research agenda meshed well with the trend in the philosophy of science and mathematics. These are also the tasks first attacked by artificial intelligence, which will be considered shortly.

Cross-fertilization of Ideas

In Chapter 8 we saw how logical positivism, quantum mechanics and relativity all gave a central theoretical role to *observation*. Investigations in all three areas showed that observation was not the simple thing it had been imagined, and it became apparent that the psychology of the observer was not irrelevant to physics. Therefore, J. Robert Oppenheimer (1904–1967), as

[2] This is because Shannon takes a message to be less informative to the extent that it is predictable. In a book of random numbers there is no redundancy; all the digits are equally likely. Hence the digits are as unpredictable as they can be.

[3] These functions are "higher" or "lower" in the sense that only "higher" animals, especially humans, possess the higher faculties, whereas all animals possess the lower.

director of the Institute for Advanced Studies, which included Einstein and von Neumann among its members, invited psychologists to visit the institute. For example, both George Miller (1920–2012) and Jerome Bruner (1915–2016) who became major contributors to cognitive science, spent a year at the IAS. Bruner had investigated the effect of the attitude and expectations of observers on their observations, work with obvious relevance to the new physics. The result was a further breakdown of disciplinary barriers, which helped the new paradigm to develop. (Miller's work will be discussed later.)

9.1.2 Birth

Cognitive Science was a Paradigm Shift

Sometimes, when pent-up frustrations with an old way of doing things have accumulated, and a number of advantages of a new way have become apparent, there will be precipitous change from the old to the new, a *phase transition* like the sudden crystalisation of a supersaturated solution. This is what seems to have happened in psychology in 1956, for in that year cognitive science was born, with its distinctively formal approach to thinking in both minds and machines. Although behavioristic methods were not abandoned, cognitive science restored mental phenomena to respectability by means of the computer model. Such a radical change of scientific methodology would later be termed a *paradigm shift* (see Sect. 11.4.4).

The First Artificial Intelligence Conference

One harbinger of the new paradigm was the first artificial intelligence conference, which was held at Dartmouth College in 1956. All the pioneers of artificial intelligence were there, including John McCarthy (1927–2011), Marvin Minsky (1927–2016), Allen Newell (1927–1992) and Herbert Simon (1916–2001), most of whom had been influenced by Ashby's *Design for a Brain*.

The Symposium on Information Theory

Newell & Simon: The Logic Theorist

A second harbinger was the "Symposium on Information Theory" held at MIT in 1956. Newell and Simon were also at this conference, at which they described the "Logic Theorist," a program that was able to prove theorems in symbolic logic from Russell and Whitehead's *Principia Mathematica*. Here was a computer accomplishing a task that would otherwise require an intelligent person, so it suggested that artificial intelligence was a real possibility, and illustrated how very precise and specific models of higher cognitive abilities could be constructed. (Newell and Simon's views are discussed in more detail in Sect. 9.3.1.)

Chomsky: Formal Linguistics

Another speaker at the symposium was Noam Chomsky (1928–), a linguist who showed how grammar could be described with complete formality, by using *term rewriting rules* similar to the Markov productions we discussed in Section 6.4.2. Chomsky's mathematical approach to linguistics superceded all others, and retains a nearly universal hegemony over the field; it's discussed in Sect. 9.2.

Miller: The Magic Number Seven

George Miller, one of the major promoters of cognitive science, also presented a paper, his famous "Magical Number Seven, Plus or Minus Two." He showed that over a wide variety of cognitive tasks the capacity of our "short-term memory" is limited to between five and nine "chunks" of information. The analogy with the "high-speed registers" of a computer (often 8 or 16 in number) was apparent and increased the attractiveness of computer models in psychology.[4]

Other Key Events of 1956

Aside from these two conferences, 1956 saw the publication of a number of other key papers signaling the birth of cognitive science. For example, anthropologist Gregory Bateson (1904–1980) applied cybernetic principles in an innovative theory of schizophrenia. *The Fundamentals of Language*, which explained language in terms of atomistic features, was published by linguists Roman Jakobson (1896–1982) and Morris Halle (1923–2018) in 1956. This year also produced *A Study of Thinking* by Bruner and his colleagues Jacqueline Goodnow (1924–2014) and George Austin (1916–2012).

von Neumann: The Computer and the Brain

Following hard on the heels of these events came further significant publications. Von Neumann wrote a set of lectures — still worth reading — that explored the similarities and differences between brains and computers, both analog and digital. Though he was too sick to deliver the lectures, they were published posthumously (von Neumann, *C&B*).

Putnam: The Mind as Software

In 1960 philosopher Hilary Putnam (1926–2016) argued that traditional mind/body problems could be solved by the computer analogy: the body was like the physical computing machine — the hardware — and the mind was like the program controlling the operation of the computer — the software. Key to this solution was the observation that software is independent of its material instantiation (as the mind is supposed to be), a property we've

[4]The most familiar example of this phenomenon is the way we hold an unfamiliar telephone number in our mind between when we look it up and when we dial it. Phone numbers also illustrate the effect of "chunking," for if part of the number is familiar, such as the area code, then it counts as only one chunk, rather than the three for an unfamiliar code. We often rehearse the contents of our short-term memory to keep from losing it, a fact which must have reminded some of the conference participants of contemporary computer memories, such as mercury delay lines and cathode-ray tube storage, both of which also required regular refreshing to retain their contents.

named *abstract formality* (Sect. 6.5.1). The resulting view, that "the mind is a program," is still hotly debated (see Sect. **??**)

Miller, Galanter & Pribram: Plans and the Structures of Behavior

The year 1960 also saw the publication of *Plans and the Structure of Behavior* by George Miller and Eugene Galanter (1924–2016), both psychologists, and Karl Pribram (1919–2015), a neuroscientist. They argued that complex structured behavior, such as Lashley had emphasized, was best explained in terms of a hierarchy of "subroutines," such as used in computer programing.[5] They also showed that many behavioral routines could be explained in terms of a common programming technique (called a *leading-decision loop* or *while-loop*[6]). One can see how artificial intelligence and cognitive psychology went hand-in-hand, for if human intelligence could be explained by a program, then (thanks to abstract formality) that same program could be executed on a computer, which as a result would behave intelligently. Conversely, if a computer program exhibited human-like intelligence, then that would be evidence that the human mind contains a similar program.

9.1.3 Growth

Hubel & Wiesel: Neural Basis of Vision

Work in all aspects of cognitive science accelerated in the 1960s and '70s. Neuroscientists David Hubel (1926–2013) and Torsten Wiesel (1924–) had been studying vision since the 1950s, and in 1962 began publishing the work that would win them the Nobel Prize in 1981. They interpreted their results as showing "geometrically" organized *feature detectors* in the brain. That is, neurons in the first stages of visual processing identified points (spots of light) in the visual field. In the second stage, points were combined to form edges and lines. The subsequent stages combined edges and lines into planar shapes, and so forth. The approach was atomistic in the logical positivist tradition, since higher-order percepts (e.g., a dog percept) were constructions of indivisible sense data (spots of light or color). Computer vision programs were organized along similar bottom-up lines: individual spots (pixels) of light were recognized, configurations of spots were assembled into edges or lines, configurations of edges and lines assembled into planar shapes, and so forth.

Early Books Pro and Con AI

[5]Indeed, Pribram was a student of Lashley; the contribution of Lashley and Pribram to connectionism is discussed later (Sect. 11.2.3).

[6]So called because it repeats some action *while* some condition holds. Miller et al., however, called it a *TOTE unit* (Test-Operate-Test-Exit).

Books reporting the first successes in artificial intelligence also appeared in the 1960s — Edward Feigenbaum (1936–) and Julian Feldman (?–) edited the collection *Computers and Thought* in 1963 and Minsky edited *Semantic Information Processing* in 1968. Not long after, the first major critique of artificial intelligence, Hubert Dreyfus's *What Computers Can't Do*, appeared; we'll consider its arguments in Section 11.3.6.

Advancement of Chomskian Linguistics

Chomsky and his students vigorously defended their formal approach to linguistics, which was presented both in Chomsky's books (*Syntactic Structures*, 1957; *Aspects of the Theory of Syntax*, 1965; *Cartesian Linguistics*, 1966; *Language and Mind*, 1972) and in anthologies, such as *The Structure of Language* (1964), edited by Jerry Fodor (1935–2017) and Jerrold Katz (1932–2002).

Influential Cognitive Science Books

The late '60s and early '70 produced several influential cognitive science books. One sign that the field was becoming recognized was the appearance of the first textbook, *Cognitive Science* (1967) by Ulric Neisser (1928–2012). Further, in 1969 Simon published his lectures, *Sciences of the Artificial*, which argued that it is possible to have empirical sciences of artificial structures, such as computer programs and human organizations, such as businesses. With Newell he published *Human Problem Solving* in 1972; it states explicitly the *physical symbol system hypothesis*, which is, in essence, the hypothesis that cognition is best understood as a calculus.

Institutional Evidence For Cognitive Science

Finally, we note evidence for the recognition of cognitive science that came in institutional form. For example, the Sloan Foundation began supporting cognitive science research in 1975. Also, in 1977 the journal *Cognitive Science* began publication, and in 1979 the Cognitive Science Society was founded. By this time the cognitive science approach was well regarded in most of the disciplines concerned with behavior: psychology, artificial intelligence, philosophy, anthropology and linguistics.

9.1.4 Characteristics of Cognitive Science

Characteristics of the Cognitive Science

Before turning to the theory of knowledge and language deriving from cognitive science, it will be worthwhile to consider the characteristics of the cognitive science approach, as explained by Gardner (*MNS*, pp. 6–7, 38–45).

Necessity of Dealing with Representation

First, in contrast to behaviorism, which rejected any reference to the mental, cognitive science took it for granted that explanations of higher cognitive functions would have to refer to plans, goals, intentions and other mental *rep-*

resentations of actual or possible states of the world. Further, it stressed the importance of structure at all levels of organization: neural, individual and social.

Thought as Digital Computation

Second, cognitive science took digital computation as the best model for the thought.[7] In part this allowed cognitive scientists to answer the objections of behaviorists: since mental representations would be modeled as specific computer data structures, and cognitive processes would be modeled as computer programs, cognitive science could promise precise testable hypotheses.

Neglect of Affective and Contextual Phenomena

Third, cognitive science concentrated on *cognitive* phenomena, such as rational thought, planning, problem solving and logic. Thus, the approach neglected, on the one hand, *affective* phenomena, such as mood and emotion, and on the other, contextual effects, such as historical and cultural factors in cognition. This betrays an "intellectualist bias" in cognitive science, which says, in effect, that cold reason is what is most characteristic of the human mind, and that nonrational emotional and social factors are simply complications of no great significance. Just as physics postulates idealized conditions, such as frictionless surfaces and perfect vacuums, to simplify analysis, so cognitive science postulated the ideal "rational agent." Divergence between the *competence* of idealized agents and the actual behavior of real people was relegated to a theoretical no-man's land called *performance*.[8]

Value of Interdisciplinary Studies

Fourth, cognitive science took the view that it was of necessity an interdisciplinary field. This was represented in the "cognitive hexagram" (Fig. 9.1), which appeared in an unpublished 1978 report (Gardner, *MNS*, p. 37). The diagram showed that cognitive science comprises six disciplines; in some cases, indicated by solid lines, interdisciplinary work was already underway; in other areas, indicated by dotted lines, it needed to be encouraged. Cognitive science was seen to be very broad in scope, encompassing structures and processes from the neural level to the social level.

The Natural Culmination of the Western Philosophical Tradition

The fifth and last characteristic of cognitive science identified by Gardner is that its agenda is essentially that of the Western philosophical tradition. Though he thinks this claim may be controversial, I hope that the preceding

[7]More specifically, as already explained, the mind was viewed as a program running on the brain as a computer.

[8]Since the hypotheses refer to idealized competence, but only performance can be observed, theories of this kind are hard to refute, since differences between prediction and observation can always be attributed to unanalyzed performance limitations. See Sect. 9.2.1 for more on the competence/performance distinction.

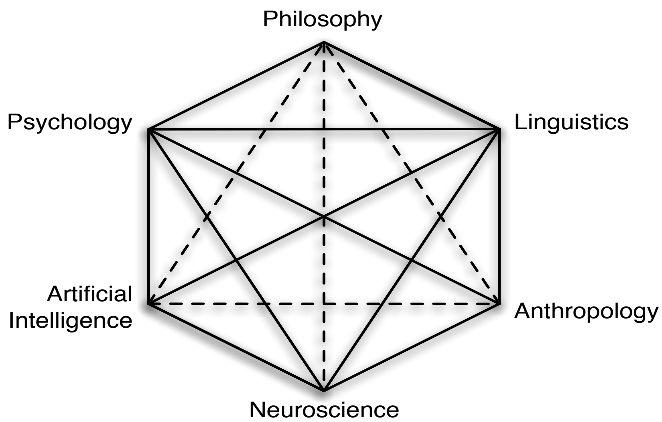


Figure 9.1: The Cognitive Hexagram. The diagram shows that cognitive science comprises six disciplines. Solid lines indicate established interdisciplinary links (as of 1978); dotted lines indicate links that needed encouragement.

chapters of *Word and Flux* have demonstrated that cognitive science was the natural, if not inevitable, culmination of a tradition extending back at least as far as Pythagoras.

Now that we've considered the historical background of cognitive science, we can look more closely at its approach to language and knowledge, as exemplified by two of its best-known theories.

* * *

The rest of this chapter and indeed most of the rest of this book exist only as a detailed outline, which follows. It may be useful as a reading list and study guide.[9]

9.2 Chomsky: Formal Linguistics

In many respects the comprehensive approach to linguistics and cognitive science pursued by Noam Chomsky and his followers is the natural culmination of the discrete approaches to knowledge representation and processing, whose development is traced in Parts I and II. As such it provides a good

[9]See the Preface for an explanation.

contrast to many of the viewpoints arising from the continuous perspective presented in Part III.

9.2.1 General Framework: Competence vs. Performance

Chomsky drew a distinction between *linguistic competence*, the abstract knowledge of a language possessed by its native speakers, and *linguistic performance*, how language is actually used. The difference is illustrated by center-embedded clauses in English. The claim is that English speakers' competence permits unlimited embedding, even though performance is much more limited (two or three levels). Chomsky argued that the linguistic competence was the proper subject matter of linguistics, but not linguistic performance, which was the concern of psychology or neuroscience (Chomsky, *ATS*, p. 3). A similar argument in cognitive science and AI concluded that the brain doesn't matter because all computers are equally powerful (equivalent to Turing machines); therefore, competence (computability) is relevant, but not performance (hardware, realtime constraints). In contrast, the alternative epistemological theories discussed in Part III have stressed the importance of performance considerations, which place strong constraints on possible theories of linguistic behavior. From this perspective, the brain and what it can do are absolutely relevant. On the other hand, these theories do not have to account for non-existent linguistic abilities, such as being able to parse arbitrarily deeply center-embedded phrases.

9.2.2 Phrase-Structured Grammars

Chomsky developed the theory of recursive phrase-structured grammars as a way of formally describing linguistic grammatical competence, in particular, unlimited center-embedding.

9.2.3 Transformational Grammars.

A transformational grammar is a system of formal rules intended to transform a sentence's *deep structure*, which directly encodes its semantic relationships, into any on of a number of *surface structures*, which represent it as a sequence of words. Thus a transformational grammar establishes a formal relation between utterances and abstract syntactic structures (often trees), which may be interpreted as expressions in a *language of thought*.

9.2.4 Government and Binding Theory.

This is a later development of transformational grammars based on two key ideas: government, which determines case relations, and binding, which determines anaphoric (pronoun) reference. [Chomsky, *The Minimalist Program* MIT Press, Cambridge MA, 1995; Chomsky, "Three Factors in Language Design," *Linguistic Inquiry* 36 (2005): 1–22.]

9.2.5 The Language Acquisition Problem

How do children acquire linguistic competence? One possibility is that they infer a formal phrase-structured grammar from samples of language to which they are exposed. Such grammatical inference is a computationally very difficult problem, and Chomsky argued that children were not exposed to enough language to support this inference (the "poverty of the stimulus"). Therefore, he argued, humans must have some sort of innate language faculty with an innate universal grammar, which can be adjusted to a particular language by early language exposure.

9.3 Symbolic Knowledge Representation

Sowa (1984) provides a nice discussion of a symbolic approach to knowledge representation and natural language processing, and shows the relevant philosophical connections. It also has a good bibliography. [J. F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, 1984.]

9.3.1 Newell and Simon: The Physical Symbol System Hypothesis

The Physical Symbol System Hypothesis: "A physical symbol system has the necessary and sufficient means for general intelligent action." In effect, the PSSH asserts that a physically-realized Turing-complete calculus is sufficient for general intelligence. [Allen Newell & Herbert A. Simon, Computer Science as Empirical Inquiry: Symbols and Search, *Communications of the ACM*, Vol. 19, No. 3 (March 1976), pp. 113–126.]

9.3.2 Schank: Conceptual Dependency Theory

Conceptual dependency theory represents meaning in terms of concepts and relations between them. All actions are defined in terms of eleven primitive actions (ACTs). For example, the meaning of "John ate a frog" refers to the primitive action INGEST, which means to take into the body. In this case INGEST has an actor (JOHN), an object (the FROG), and an associated DIRECTION (another primitive) from an unknown location Y into MOUTH. It also has a default INSTRUMENT, which is another primitive action, MOVE, with actor JOHN, object HAND, and DIRECTION from the same location Y into MOUTH. In this way, meanings (conceptual dependencies) can be represented, including their expected preconditions and consequences. For example, if a story is translated into its conceptual dependencies, then reasonable inferences can be drawn even if they were not expressed explicitly in the story. [Roger G. Schank, *Conceptual Information Processing*, North-Holland, Amsterdam, 1975. See esp. ch. 3, §§3.1–3.5 (pp. 22–49).]

a. Introduction
b. Instruments and Causation
c. Conceptual Roles and Conceptual Rules
d. The Primitive Actions
e. States

9.3.3 Abelson: Scripts

A script is an abstract structure representing a particular scenario. The classic example is the RESTAURANT SCRIPT, which describes stereotypical aspects of eating a meal at a restaurant. It has slots with default fillers for the various parts of the script, for example, that you order from the menu, that you eat the food, that you pay your bill, etc. These are things that will be assumed unless something is stated to the contrary. It thus defines the background assumptions of the scenario. The problem is that real-life scenarios have a seemingly infinite set of unarticulated background assumptions. For example, a restaurant script would not normally include the fact that you don't eat the waiter, but that is part of the implicit and usually unarticulated background. In a script, all the background assumptions should be made explicit, but that is impossible. [Robert P. Abelson, Psychological Status of the Script Concept, *American Psychologist*, Vol. 36, No. 7 (July 1981), pp. 715–729.]

a. Scripts in Artificial Intelligence
b. Scripts as Cognitive Structures
c. Scripts as Performative Structures
d. Scripts Related to Other Theoretical Constructs
e. Metascripts and Other Generalizations
f. Conclusions

9.3.4 Expert Systems vs. Expert Behavior

a. Expert Systems. Expert systems attempt to express expertise as collections of rules with an IF-THEN structure. *Knowledge engineers* interview experts and take protocols to extract the (alleged) rules that the experts follow.

b. Dreyfus Model of Skill Acquisition. Hubert and Stuart Dreyfus classified skillful human behavior into five levels: *novice, competence, proficiency, expertise*, and *mastery*. They discovered that non-contextual rule following (similar to expert systems) is characteristic of the advanced novice level, and that true experts have moved beyond rule use (although applicable rules may be cited post hoc as justification). At the level of proficiency and above, decision making is guided by intuition. [Dreyfus, Hubert & Dreyfus, Stuart (1986), *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, Oxford, U.K.: Blackwell. The model has been criticized by Gobet, F. and Chassy, P. (2009), Expertise and intuition: A tale of three theories, *Minds and Machines*, 19, 151–180.]

9.3.5 The Frame Problem

The *frame problem* arises from applying formal logic in robot reasoning. Which propositions change and which do not change as a result of an action? More generally, how is it possible to reason about the physical world? It is ultimately an issue of relevance and our background understanding of the world. [Zenon Pylyshyn, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Ablex Publishing, 1987.]

9.3.6 The Language of Thought Hypothesis

It has been claimed that there must be a *language of thought* ("mentalese") because "it is the only game in town." That is, to account for the structural richness of knowledge, it must be represented in language-like discrete structures. Therefore, all thought is at some level discursive. [Fodor, Jerry A., *The Language Of Thought*, Harvard Univ. P., 1975. Fodor, Jerry A. and Zenon W. Pylyshyn (1988), Connectionism and Cognitive Architecture: A Critical Analysis, in S. Pinker and J. Mehler, eds., *Connections and Symbols*, Cambridge, Massachusetts: MIT Press (A *Cognition* Special Issue).]

9.3.7 Conclusions

Symbolic artificial intelligence and symbolic cognitive science both assume that knowledge can be represented as discrete atoms, representing categories, connected by discrete relationships, with semantics similar to propositions in formal logic. Cognition, then, is a process of transforming these structures by means of discrete formal rules. Or, to put it in other terms, cognition can be described by a calculus. The inadequacies of this hypothesis are revealed by the Dreyfus model, the frame problem, the language acquisition problem, performance issues, etc. All these point to the need for an alternative explanation of knowledge and cognition, that is, to a different understanding of epistemology.

Part III

Alternative Views of Knowledge

Chapter 10

Flux and Strife

10.1 Introduction

We return again to ancient Greek philosophy, and you may well ask why. We have seen the ubiquity of the idea of a calculus in western philosophy's view of knowledge, beginning with Pythagoras' description of nature as formal arrangements of tokens and, through the medium of Plato's and Aristotle's epistemological theories, culminating in modern computational theories of the mind. This view of knowledge has formed so integral a part of the philosophical background that it has rarely been questioned.

Nevertheless, there have always been dissenters from the traditional view of knowledge; indeed the earliest surviving fragment of Greek philosophical prose is just such a dissent (see p. 360, n. 9). Throughout the centuries there have been philosophers and others who have realized that there is more to knowledge than that which can be put into discrete symbolic structures; it will be worthwhile to familiarize ourselves with their observations and insights. We will see that, complementing the theoretical limitations of the discrete discussed in the last chapter, there are also philosophical, psychological and biological reasons for questioning the old view. Part III is devoted to these reasons.

10.2 Heraclitus: Flux and Logos

Nought may endure but Mutability.

— Shelly, *Mutability*

10.2.1 Background

Heraclitus: fl. 500 BCE

Little is known with any certainty about the life of Heraclitus. It is generally agreed that he lived in Ephesus (Asia Minor) around 500 BCE, but there is even disagreement about this simple fact. What is known with confidence is the profound influence that Heraclitus exerted on all his successors. There is some evidence that Plato may have been a student of the Heraclitean philosopher Cratylus, and Plato's dialogues contain numerous — mostly ironical — references to Heraclitean doctrine. Aristotle also has much to say about Heraclitus, and the Stoics adopted him as their patron saint (distorting his ideas in the process). Indeed, much of western philosophy can be described as a search for stable rational structures (*logoi*) in the Heraclitean flux.

Textual Problems

Heraclitus may have written a book and it may have been called *On Nature*. This was the position of ancient authorities and has been maintained by some modern authors, such as Kahn, on the basis of the thematic unity of the fragments that remain. (In fact, Kahn believes that this book survived intact until 200 CE and perhaps as late as the fifth century.) In contrast, other scholars, such as Diels and Kirk, see only a hodge-podge of aphorisms that might have been collected after the fact into a book.

In any case, all we have now are some 130 fragments, many less than a line long. Furthermore, since there were no quotation marks in ancient times, it is frequently unclear whether an ancient author was quoting or paraphrasing Heraclitus, and it is difficult to tell where a quotation begins and ends. As a result, arguments about the authenticity and extent of the fragments often hinge on claims about the words and forms Heraclitus would have "likely" used. These arguments are complicated by the corrupt state of many of the surviving manuscripts and by the fact that the sources of many of the fragments are authors with their own axes to grind. Indeed, the fragments have provided a kind of Rorschach test for philosophers of almost every persuasion (including the Stoics, the Neoplatonists, the Christian church fathers, Hegel, Nietzsche, Marx and Lenin).[1]

10.2.2 Heraclitus' Style

A wise man is not likely to talk nonsense.

— Plato, *Theaetetus*, 152b1

Heraclitus The Obscure

Heraclitus is not easy to understand. It might seem that this is because we have only fragments of his writings, pithy bits and pieces that other writers were inclined to quote. This is no doubt part of the problem, but even in antiquity when Heraclitus' entire book was extant, he was known as *The Obscure* and *The Riddler*. Socrates, Plato and Aristotle all complained that he was hard to understand. There is a story that Euripides gave a copy of Heraclitus' book to Socrates and asked his opinion of it. Socrates answered,

[1] The collection of the fragments of Heraclitus' book that I have found most useful is Kahn (*ATH*), which includes the fragments in Greek and English (organized in accord with Kahn's interpretation, with which I do not entirely agree), extensive commentary as well as helpful notes on vocabulary, text and context. Second most useful has been Kirk, Raven & Schofield (*Presoc.*, Ch. VI), which presents the fragments topically in Greek and English with commentary, and perhaps represents the consensus interpretation of Heraclitus. One of the earliest collections of the Greek text is Bywater (*Her.*), which also attempts a topical organization. English translations following the Bywater organization can be found in Burnet (*EGP*, pp. 132–141), Nahm (*SEGP*, pp. 89–96) and Patrick (*Her.*) (the latter bound with Bywater, *Her.*, in Richards, *Her.*); both Bywater (*Her.*) and Patrick (*Her.*) give information on the contexts in which the fragments were quoted. For many years the definitive collection of fragments has been Diels & Kranz (*Frag.*, 22B), although this suffers from the lack of an organizing interpretation (thus reflecting Diels' view that Heraclitus had no organized doctrine); the translations in Freeman (*APSP*, pp. 24–34), which follows the Diels organization, are not very helpful. The Diels order is also followed by Jones (*Hipp.*), which provides the Greek and English on facing pages.

There are several collections of the fragments in English translation. Of these, Barnes (*EGP*, Ch. 8) is especially useful in showing the contexts of the fragments. The fragments are organized topically in Wheelwright (*Her.*), although I find his translations misleading. Wheelwright (*Pres.*, Ch. 3) contains the same translations, as well as translations of many ancient commentaries on Heraclitus. Several other books quote large numbers of the fragments *en passant* (which gives them a topical organization), including Guthrie (*HGP*, Vol. I, Ch. VII), which is definitive, Robinson (*IEGP*, Ch. 5), which includes many of the fragments, and Hussey (*Presoc.*, Ch. 3), which is an informative, brief introduction.

> What I understand is excellent, as is no doubt the part I don't understand. But it takes a Delian diver to get to the bottom of it. (Diogenes Laertius, ii.22)

There were also some in antiquity who thought that Heraclitus was intentionally obscure, since he wanted to be understood only by the best and brightest, who would devote the necessary effort:

> He dedicated and placed [his book] in the temple of Artemis, as some say, having purposely written it rather obscurely so that only the able should have access to it, and it should not be easily despised by the people. (Diogenes Laertius, ix.6; tr. after Kirk, Raven & Schofield)

More likely these stories were invented to explain the difficulty of his ideas. Although may seem that he didn't want to be understood, I will argue that Heraclitus' style is just what was required to express his message.

Gnomic Style and Recognition

In modern times his style has been described as oracular, aphoristic and *gnomic*, because he wrote in *gnomes*: aphorisms or maxims. As we will see, many of Heraclitus' truths cannot be said, but must be shown, and this is what a gnome does; recall the discussion of the related word *gnomon* (p. 23). A gnome is understood by a process of recognition (*gignōskō*) related to the recognition of a familiar face or situation. This is the most important kind of cognition for Heraclitus (Kahn, *ATH*, pp. 104–104, 171), and I will have much more to say about it in later chapters.

Stretching Language

Languages evolve to fit the needs of the communities that use them. Therefore, when it's necessary to say something very new, the available language may prove inadequate, which is the situation in which Heraclitus and the other early philosophers found themselves. Greek had not yet developed the abstract vocabulary that the European languages have inherited and on which we have come to depend. When new thoughts are to be expressed, the language must be stretched; since Heraclitus' thoughts were very original, we must expect as much from his use of language.[2]

Creating New Categories

One way in which language is extended is by creating new categories along with expressions for evoking them. This is accomplished by establishing new

[2] Hussey (*Presoc.*, pp. 34, 59) claims that Heraclitus and the early Wittgenstein were both motivated to use language in innovative ways by recent developments: the alphabet and explicitly formulated general laws in the case of Heraclitus, symbolic logic in the case of Wittgenstein. Both developments promised the possibility of logoi that would depict reality more accurately than common speech. (See also Section 11.1.2.)

patterns of association between the experiences, ideas and words already shared by the language community. It is the mark of a great artist, poet or philosopher to be able to create the perceptual experiences (such as texts and other works of art) that will resculpt the cognitive landscape in accord with his or her vision. And so we find that Heraclitus' style has much in common with that of the Seven Wise Men and poets such as Pindar and Aeschylus (Kahn, *ATH*, pp. 6–7).

Symbols

There are several ways to remodel the field of associations; one is to reshape a recognizable thought or experience that already has in part the desired form. By modifying old relations and creating new ones, the familiar may become a *symbol* for the novel. Symbols are most commonly needed when the language lacks adequate means to express the new idea.[3]

Metaphor and Analogy

As a symbol becomes established it may be possible to transfer to it forms of association patterned after common, shared experiences; this can be accomplished by simile, metaphor and other artistic devices. If the resources of the language are sufficient to express the relationships, then they may also be established by formal analogy.[4]

Words

The are many ways the artist may seize and manipulate the field of associations. In verbal communication words are the most accessible handles; they may be pulled, pushed and related syntactically in order to effect changes in the associative field. To this end the poet may exploit denotations, connotations, forms, sounds and etymologies. Further, etymologies, like myths, need not even be true to accomplish their communicative purposes; they are merely tools for sculpting the associative landscape. (This may explain early philosophy's fascination with speculative etymology; see for example Plato's *Cratylus* and the quotation from it on p. 372).

Reading Heraclitus

The surviving fragments of Heraclitus show him to be a master of these literary devices, and Kahn (*ATH*, pp. 87–95) has argued convincingly that understanding Heraclitus' thought requires sensitivity to them. In other

---

[3]Thus it is nor surprise that the use of symbols is frequent in Heraclitus and a characteristic of archaic logic (Prier, *AL*, pp. 5–8).

[4]'Metaphor' comes from a Greek word meaning 'to transfer' and is thus a device for transferring associations from one concept to another. In an analogy, by contrast, two verbal formulas are set down beside each other so that an exact correspondence can be established (*analogos* = according to a *logos*; the original meaning of *analogia* is proportion, which shows its formal connotation). Plotinus had this to say on Heraclitus' style: "He seems to speak in similes, careless of making his meaning clear, perhaps because in his view we ought to seek within ourselves as he himself has successfully sought." (Plotinus *Enneads* iv.8, p. 468; tr. Guthrie)

words, we cannot read the fragments like propositions in a logical argument; we must read them as literature — even poetry. Two of Heraclitus' means of expression deserve special mention, because they illustrate important mechanisms of cognition; Kahn calls them *linguistic density* and *resonance*.

Linguistic Density

Linguistic density refers to the strength and richness of associations between a word and other thoughts or experiences; we may call the totality of these the *manifold* controlled by the word.[5] Some words have denser manifolds than others, but in all cases the manifold is a resource that the writer or speaker exploits to artistic ends.

Apparent Lexical Ambiguity

Unfortunately it is difficult to preserve in translation a writer's use of density. Since languages have evolved to meet the needs of their communities, the manifold controlled by a single word in one language may be controlled by several words in another. From the perspective of the second language the word appears to be ambiguous, although it's rarely experienced as ambiguous by speakers of the first language (Chukovsky, *AT*, pp. 55–59).

Logos

We have already seen an example of apparent ambiguity in ancient Greek *logos*. The oldest meaning is 'speech', but the demands imposed by the invention of writing and more abstract thought extended its manifold to include the uses of the English words 'account', 'record', 'reason', 'reasoning' etc. (see pp. 21 and 361).

Preserving Density in Translation

Accurate translation requires picking a word or expression that controls a similar manifold to the original word; when this is impossible there may be no choice but to retain the original (as I have done with *logos*). The only alternative is to pick one of the possible translations, but doing so presumes that only part of the original word's manifold was relevant to the writer's purpose and thus imposes a limited interpretation on the text. In some cases a limited interpretation may be justified, but when we are dealing with a literary master it's safer to assume that he or she was drawing on all the resources of the word's manifold.

Apparent Syntactic Ambiguity

Linguistic density is not just a property of individual words; it's also a property of more complex expressions, including phrases and entire utterances, since they also have their manifolds. Here too there may be apparent ambiguity since the manifold controlled by one utterance may nearly coincide with that controlled by two or more distinct utterances. Picking one of these, in translation or interpretation, may cut the original manifold in a

[5]These manifolds are related to the *semantic fields* discussed by Nalimov (*ILL*), but include the connotations as well as the denotations of words; see Sec. 11.2.2.

way not intended by the author. If we believe this "ambiguity" is intentional — if we believe it is a tool the author is using to remodel our understanding — then we must try to preserve it in translation.

Overcoming the Limitations of Linear Expression

The manifold that Heraclitus wishes to communicate is a dense, multidimensional complex of ideas and experiences, but speech and writing are linear. It would be very inefficient to try to communicate with precision the multidimensional complexity of such a manifold by a simple, literal linear description.[6] It is much more efficient to use the associative fields of the words to establish connections that reach across the linear structure of the utterance: repeating a word recalls its earlier use; using a related word recalls the associations of the former; even a phonetic or lexical similarity may be used to establish connections.

Resonance

All these devices make use of *resonance*: situations in which multiple occurrences of a word, idea, image, symbol or sound resonate together to enrich the connections in the manifold. Seen in this light, word play is not a triviality to be ignored or relegated to footnotes; it is a resource necessary to the author to the extent that the manifold is complex in structure and rich in texture. We should not be surprised to discover all kinds of resonance in Heraclitus, and we must try to be aware of it if we want him to speak to us.

10.2.3 Interpreting Heraclitus

> He is much too big for our formulas.
>
> — Burnet (*GPI*, p. 57)

Communication Against a Background

I have observed above that speakers and writers accomplish their communicative purposes by manipulating the field of associations of their audiences.

[6]Guthrie (*HGP*, Vol. I, pp. 453–454) notes the futility of a linear presentation of Heraclitus' doctrine: "what is wanted is a simultaneous advance on several fronts at once." We may compare the difficulties Wittgenstein had presenting his later ideas sequentially: "my thoughts were soon crippled if I tried to force them on in any single direction against their natural inclination . . . the very nature of the investigation . . . compels us to travel over a wide field of thought criss-cross in every direction" (Wittgenstein, *PI*, p. ix). (See Section 11.1.2 and also compare p. 354.) Similarly, Heraclitus' mode of expression produces meaning by "appositional clusters," in which a basic fact is stated and then elaborated factually and emotionally (Prier, *AL*, pp. 10–16; Thornton & Thornton, *T&S*, pp. 75–88). As in music, there is statement, development and recapitulation of the themes, resulting in a progressive accumulation of meaning.

To be able to do this effectively they must have some intuitive understanding of the shape of the common cognitive landscape, an understanding which comes largely from sharing with the audience a "background" of common beliefs, expectations, hopes, fears, prejudices etc. — in brief, a *world*. Heraclitus lived in his world and we live in ours. They are alike in some ways, since we are all humans and certain basic concerns are common to all people; that is why the classics still speak to us. In other respects, the world of a fifth century BCE Greek living on the coast of Asia Minor was very different from ours. Can we hope to understand Heraclitus?

Necessity of Interpretation

To understand Heraclitus as he was understood by fifth century Greeks we would have to *be* fifth century Greeks, and that is clearly impossible; we cannot change the fact that we are late twentieth century readers. Further, we hope that Heraclitus will give us insights into *our* concerns and problems; we are not so interested in those peculiar to his contemporaries. Thus we must *interpret* Heraclitus' writings in the context of our own world.

Interpretive Frame May Be Chosen

When we read a text it is always in the context of some interpretive framework, otherwise it would be meaningless to us, since the text generates its meaning by operating on the framework. The interpretive frame is partly a matter of choice; we can change our viewpoint, bring a different frame to a text and read it in a different way. But this very flexibility makes the choice of frame an *issue*: Which interpretation is correct?

Evaluation of Interpretive Frame

One test of an interpretation is to see if it is consistent with the interpretation of other, contemporary texts, since they were presumably aimed at similar audiences. We also expect coherence between the interpretation of a text and those of earlier and later texts, since a later text is often a response or reaction to an earlier. Thus we may justify one interpretive frame by an appeal to others, but how do we justify those others? The argument looks circular.

All Evidence Interpreted

An apparent escape is to evaluate an interpretation on the basis of its consistency with nontextual evidence, such as archaeological evidence, but this ignores the fact that archeological evidence must also be interpreted. The full meaning of an artifact is a function of the background of needs, concerns and expectations of the society that created it; because it is not our society we have no direct understanding of this meaning. Further, the interpretation of archaeological evidence is often based on textual evidence, and so the interpretive circle is closed.[7] This is the *hermeneutic circle* (from

[7] It is important to keep in mind that even so basic a tool as a dictionary is based on

hermēneus = interpreter), a central topic of Section 11.3.

Hermeneutic Circle Inescapable

Texts and other evidence are understood in the context of an interpretive frame, but an interpretive frame is justified by the sense it makes of the evidence. The hermeneutic circle is inescapable. The best we can hope to do is to assess how well *all* our interpretations — textual, linguistic, archaeological — hang together, how convincing a story they tell.

Invisibility of Familiar Frame

A comprehensive interpretive frame must be evaluated from within, for there is no outside. If we think we are outside, it simply means that we are using a frame that is so familiar, so comfortable, that we do not notice its presence. Like the frame of our eyeglasses, it is so close that we see through it and lose sight of it; it's invisible by its familiarity. This is a dangerous situation, since it means we are making interpretive assumptions without being aware of them. (Indeed, the goal of Parts I and II of this book has been to expose the Pythagorean frame through which we are accustomed to understand knowledge and cognition, so that we may question that frame and explore alternative perspectives.) It is better to acknowledge that we are caught in the hermeneutic circle; then we can make our interpretive choices consciously.

Alternate Frames Seem Unreasonable

To reiterate, we cannot evaluate a frame from "outside," since there is no outside; we can only evaluate a frame from within a frame, either itself or another. One of the impediments to choosing a new interpretive frame is that from the perspective of one, another usually looks inconsistent (since it's incoherent with the first). Therefore, refining a familiar frame usually seems more sensible than adopting a new one. (Of course, a frame may develop so many internal problems that it demands its own abandonment; this has to some extent happened to the Pythagorean frame, which has revealed its own limitations; recall Chapter 7.)

Necessity of Bold Interpretation

Thus the only way to evaluate a new frame fairly is to enter it boldly and to see how it works from the inside. This approach is all the more necessary for interpreting Heraclitus, since the fragmentary text does not constrain the possible interpretations to the extent a complete text would. If we do not venture widely we will be unlikely to recover the original meaning of Heraclitus' text or to understand his thought.

Interpretations

My interpretation of Heraclitus, so far as I can tell, has been influenced most strongly by Popper (*C&R*, Ch. 5) and Popper (*OS&E*, Ch. 2), and to a lesser extent by Hussey (*Presoc.*, Ch. 3) and Kahn (*ATH*). Prier (*AL*,

---

interpretation of textual, linguistic and archaeological evidence.

pp. 57–61) says most interpretations of Heraclitus fall into two categories: "English," which has a scientific orientation (e.g., Guthrie), and "German," which has an idealist (e.g., Hegel). By this classification, my interpretation is more "English," but without the usual accompanying dualism.[8]

10.2.4 The Way of Inquiry

Difficulty of Understanding the Logos

According to ancient authorities, Heraclitus' *On Nature* began:[9]

> (1) Although this *logos* holds always men are uncomprehending, both before they have heard it and once they have heard it. For although all things come to be according to this *logos*, men seem inexperienced when they experience such words and works as I set forth, separating each according to its nature and telling how it is. But other men lose sight of what they do awake, just as they lose thought of what they do asleep.

The full significance of this is by no means apparent, and its explication is the purpose of the rest of this section, as it probably was of Heraclitus' book.

Example of Ambiguity

As though to warn us of the difficulties ahead, Heraclitus puts a syntactic ambiguity in his first line; as Aristotle noted (*Rhetoric* 1407b14–18), the meaning is different if one puts a comma before or after 'always'. The latter

[8]Various authors have treated different ideas as "central" to Heraclitus' doctrine: *universal fire:* [Diogenes Laertius, ix]; *universal flux:* Zeller (*OHGP*), Popper (*OS&E*), Popper (*C&R*), Wheelwright (*Her.*) and Robinson (*IEGP*); *unity of opposites:* Burnet (*GPI*), Burnet (*EGP*) and Hussey (*Presoc.*); *lawful change:* Guthrie (*HGP*, Vol. I) and Kirk, Raven & Schofield (*Presoc.*); *unity of human condition and nature:* Kahn (*ATH*). I don't think this issue is very important, however, since these ideas are all melodic lines in the Heraclitean counterpoint.

[9]This is probably the oldest surviving fragment of Greek prose that is more than one sentence long (Kahn, *ATH*, p. 96). Unless otherwise noted, the translations in this section are my own. My reason for doing this is to try to better preserve (to an even greater extent than Kahn) the linguistic density and resonance of Heraclitus' text. In my translations I have of course leaned heavily on the experts, especially Kahn (*ATH*) and Kirk, Raven & Schofield (*Presoc.*). See Moser (GEB) for an informative discussion of the problems of translating a book that depends heavily on word play and resonance. Each Heraclitean fragments is given a number in this section, according to my topical ordering. The table on p. 382 shows the corrspondence between these numbers and their numbers in section 22 B of Diels & Kranz (*Frag.*); Bywater (*Her.*), which attempts a topical organization, and Kirk, Raven & Schofield (*Presoc.*), which is perhaps the most accessible source.

stresses the claim that people do not now comprehend the *eternally* valid logos; the former that although the logos is a fact, people *never* comprehend it. Both interpretations seem to be consistent with Heraclitus' thought, and in the 2300 years since Aristotle's complaint different scholars have placed the comma differently. However, since this paragraph seems to be very carefully crafted, it is more likely that the ambiguity is intentional, and that Heraclitus is trying to convey *both* interpretations at once (Kahn, *ATH*, pp. 92–95, 97–98); perhaps he even intends the perpetual lack of comprehension to be an example of the eternal logos. The issue is not one to be addressed now; it is presented as an illustration of Heraclitus' polyphonic writing. To understand his theme, we must be sensitive to both the melody and the harmony.

Deceptiveness of Appearances

Heraclitus' goal is to present the logos of the universe; for now we may interpret this either as the underlying principle of the universe or as an account of that principle — the ambiguity is no doubt intentional.[10] However, Heraclitus has warned us that this principle is difficult to comprehend, and that although we are constantly in contact with it, we do not recognize it for what it is.

Relativity of Subject

Part of the difficulty of understanding the logos is that we view it through our categories, through oppositions of A and not-A. Many of the fragments warn us of the dangers of a rigid categorical system. For example:

(2) Sea is the purest and foulest water: for fish drinkable and saving, but for men undrinkable and destroying.

(3) Asses would rather have sweepings than gold.

(4) Swine enjoy mud more then pure water.

(5) Swine wash in mire, and flocks of birds in dust or ashes.

Thus, being pure or foul is not an objective property of the sea; the category to which it belongs depends on whether you are a human or a fish; the categories are *subjective* rather than *objective*.

Relativity of Purpose

[10] Recall that *logos* has a constellation of meanings, which in the ancient Greek of Heraclitus' time include: word, utterance, report, account, reputation, measure, proportion, analogy, correspondence, relation, cause, a reason, reasoning, the faculty of reason, argument, principle, definition. See also the discussion on p. 21; for additional information on Heraclitus' use of *logos* see Guthrie (*HGP*, Vol. I, pp. 419–424), Hussey (*Presoc.*, pp. 40–41) and Peters (*GPT*, pp. 110–111). Kerferd (*SM*, pp. 83–84) shows how linguistic density results from the simultaneous application of *logos* to language, thought and the world.

There are other sources of relativity in opposition. For example, activities that are bad and undesirable in some situations may be good and desirable in others. For example, ancient medicine was often indistinguishable from torture, except in regard to its purpose:

(6) Doctors who cut, burn and wretchedly torture the sick in every way complain that they do not receive the reward they deserve. (tr. after Kahn)

We may call this *relativity of purpose*.

Relativity of Perspective

Another source of difficulties is *relativity of perspective*. For example, some scholars interpret

(7) The way up and down is one and the same.

to mean that the same path is "the way up" to someone at the bottom of the hill and "the way down" to someone at the top. (Another interpretation is discussed on p. 366.)

Relativity of Aspect

Another problem with naming things may be called *relativity of aspect*:

(8) The way of the writing, straight and crooked, is one and the same.

In other words, we may call the writing straight since it goes in lines from left to right, but also crooked since it twists and curves; in fact it is both, but the categories apply to different aspects of the writing.[11]

Context

Later (Section 10.2.5) we will see other examples of how opposing categories misrepresent the world; for now the relevant point is that *categories are context-dependent*, where we include in the context subject, purpose, perspective, aspect and (as we will see later) time.[12] Since we are always in a context, these paradoxes show that we cannot take our perceptions at their face value and that we must treat with circumspection the words we use.

Much Learning vs. Wisdom

How then can we understand the logos of the universe? Wide experience certainly forms a necessary basis:[13]

---

[11]Some read 'carding wheels' (γναφέων) instead of 'writing' (γραφέων), in which case the fragment probably refers to their simultaneous linear and cyclic motion. Under either reading the point is the same.

[12]See also Kirk, Raven & Schofield (*Presoc.*, p. 189) and Guthrie (*HGP*, Vol. I, pp. 445–446).

[13]'Wisdom-loving' = *philosophos*; if the quotation is genuine, this is the earliest extant use of this word Kahn (*ATH*, pp. 105, 308), and so its etymology would be especially salient. It may also allude to the Seven Wise Men (*Sophoi*). Finally note the 'many' could refer to things or people.

(9) Wisdom-loving men need indeed to be inquirers of very many.

This is because recognition works through our most reliable source of knowledge, the senses:

(10) Of whatever there's sight, hearing, perceiving: these I prefer.

But we've seen that a naive appeal to perception will lead us astray unless it is tempered with understanding:

(11) Bad witnesses are eyes and ears for men if they have barbarian spirits (*animae*).[14]

Since most people have "barbarian spirits" they cannot rightly interpret the logos, although they are in constant contact with it:

(12) The uncomprehending, listening, seem deaf; the saying bears witness to them: Present, they are absent.

Their knowledge is superficial; they think they understand, but they do not:

(13) For the many do not think about things in encountering them, nor recognize when perceiving, but to themselves seem to do so.

Learning is useless unless it can be integrated into a recognition of the truth. Even the most learned have only a superficial understanding:

(14) Much learning does not teach sense, for otherwise it would have taught Hesiod and Pythagoras, and again Xenophanes and Hecataeus.

For they are "caught in a web of words":

(15) Pythagoras, son of Mnesarchos, practiced inquiry most of all men, and, choosing from these writings, made a wisdom of his own — much learning, evil practice.

This juggling of words that are only superficially meaningful is an "evil practice," of which Pythagoras is the chief practitioner.

Necessity of Deep Inquiry

If the accepted way of inquiry is futile, then how can we hope to comprehend the logos? No one has recognized that the true account must be very different from the usual view of things:

[14]For *anima* see p. 379.

(16) Of those whose logoi I have heard, none has come so far as to recognize that what is wise is separated from all.

The reason is that

(17) Nature is wont to lie hidden.

Therefore,

(18) If one does not expect the unexpected, one will not find it out, for it can't be found out and is impenetrable.

Or, as Pasteur said over two millennia later, "Where observation is concerned, chance favors only the prepared mind."[15]

To discover the logos of the universe we will have to look deeply, and we must be prepared for an account that goes against common sense and that appears paradoxical. Indeed, since the familiar words and categories obscure the true nature of things, we must expect it to be difficult to put this logos in words (itself a paradox). The words we use, like the oracles of Apollo, will point to the truth rather than saying it directly:

(19) The lord whose oracle is in Delphi neither declares nor hides, but indicates.

Such is the nature of the deepest truths:[16]

(20) "The Sibyl with raving mouth," according to Heraclitus, "uttering things mirthless and unadorned and unperfumed," reaches over a thousand years with her voice through the god. (Plutarch *Why the Pythia No Longer Prophesies in Verse* 6.397A)

Seeking the Common

We may think we know many things, but as long as our knowledge is in terms of superficial, context-dependent categories, it is a false understanding, for wisdom is understanding the basic rule of the universe:

(21) The wise is one: to know the gnome (judgement, maxim), how all things are steered through all.[17]

---

[15] *Dans les champs de l'observation le hasard ne favorise que les esprits préparés.* (Address at the University of Lille, Dec. 7, 1854)

[16] There is doubt about the extent of the quotation; see Kahn (*ATH*, pp. 124–125).

(Recall the meaning of *gnome*, p. 354.) Therefore, if we want to achieve more than "private thoughts," we must seek what is common to all contexts:

(22) Therefore it is necessary to follow the common; but although the *logos* is common the many live as though they have a private thinking. (tr. after Kirk, Raven & Schofield)

Our first step toward objectivity ("the common") will come from a better understanding of categorical knowledge (Section 10.2.5).

"The Common"

But what "is the common"?

(23) Thinking is common to all.

(24) Heraclitus says that for the waking the world is one and common, but the sleeping turn aside each into a private one. (Plutarch)

(25) Those speaking with sense (*xun noōi*) need to rely on what is common (*tōi xunōi*) to all, as a city on its customs (*tōi nomōi*) — and with much more reliance. For all human customs are nourished by one, the divine. For it dominates as much as it will and suffices for all — and exceeds.

Notice how in Fr. 25 Heraclitus uses word play to lend unity to his thought. The Greek *nomos* means law as well as custom; perhaps 'common law' is a good translation, since it reflects the near coincidence of law and custom in ancient society (see also Kahn, *ATH*, p. 118; Kerferd, *SM*, 112–113; Popper, *OS&E*, p. 17–18). The fragment seems to mean this: Within the city, speaking with understanding depends on a common background of beliefs and customs, otherwise there is no basis for communication or common understanding; but this is still a parochial understanding. Speaking with *full* understanding requires recognizing what is common to all (people or things? — the word is ambiguous), since all particular customs have a basis in one (custom or thing — ambiguously). The thought forshadows Heidegger (Section 11.3.5).

[17] The text is uncertain (Kahn, *ATH*, pp. 170–171; Kirk, Raven & Schofield, *Presoc.*, p. 202). Another reading is: "be sure of the gnome by which it pilots all through all." Under either reading, wisdom is recognizing the guiding principle of the universe. It is interesting that the word for 'is steered', *kubernatai*, comes from the same root as our *cybernetics*, the study of control processes in natural and artificial systems. Thus the fragment says, in effect, that wisdom comes from the recognition that the universe is a cybernetic system, and from understanding the principle governing the mutual interactions of everything in the universe. See also Section 10.2.7.

10.2.5 Oppositions in Continua

Mutual Determination

Heraclitus shows that the identity of each category depends on its opposite, and therefore that there is an underlying unity in oppositions. For example, recognizing the category *justice* requires acquaintance with *injustice*:

(26) The name of Justice they would not know, if these things were not.

Apparently contradicting Thales,[18] who said, "The sweetest thing is to obtain what you desire," Heraclitus asserts:

(27) It is not better for men to have all they wish. It is sickness that makes health sweet; evil good; hunger glut; weariness rest.

There is one unity of phenomena out of which are separated the categories just/unjust, sick/healthy, etc.

It is important to recognize that the theory of unity in opposition applies to itself, since it makes use of the categories *unity* and *opposition*, which are opposites. Heraclitus seems to have recognized this self-applicability (Hussey, *Presoc.*, p. 45):

(28) Comprehensions: wholes and not wholes, convergent divergent, concordant discordant, one from all and all from one.

Each of the pairs can be taken together as a whole — a comprehension — but they are also not wholes since the opposites can be taken separately.

Complementary Changes

An important characteristic of many oppositions is that they are complementary: more of one is less of the other. We have already seen this for sickness/health, hunger/glut etc. This fragment provides physical examples:

(29) Cold warms up, warm cools off, moist parches, dry dampens. (tr. Kahn)

In general the increase of one is the decrease of the other, which is another interpretation of "The way up and down is one and the same" (Fr. 7). Again, the underlying unity is apparent: the same change is both increase and decrease. This unity provides a deeper understanding of natural phenomena. For example, in Hesiod's *Theogony* (744–757) daytime and nighttime result

[18]Thales (fl. 585 BCE): one of the Seven Wise Men, and often considered the first philosopher in the Western tradition.

from the alternating dominance of Day and Night, which are independent forces (goddesses).[19] However, astronomical observation eventually showed that the lengths of the day and the night add to a constant, twenty-four hours, and therefore that the increase of day is the decrease of night, and vice versa. The unity underlying the opposites day and night was even given a name, night-and-day (*nuchthēmeron*; Kahn, *ATH*, p. 109). This seems to be the point of:

(30) The teacher of most is Hesiod; they are sure he knows most, who
did not recognize Day and Night — for they are one.

The opposites are superficial divisions of a more fundamental continuum.

10.2.6 Universal Flux

(31) The changing thing rests. (*Metaballon anapauetai.*)

Since the increase of one force leads to a decrease of its complementary opposite, we may view the opposition as a "war" in which either there is a balance of the forces, or in which first one side and then the other prevails. Heraclitus illustrates the case of balanced change with a famous metaphor:[20] *Equilibrium*

(32) On those in streams
self same stepping
other and other waters flow.

The stream, if it is a stable, persistent object, must have a steady flow into and out of any particular place. The image of the stream displays a typical Heraclitean opposition. In one sense the stream is *not* one "self same" thing, since it changes from moment to moment; it is never the same as itself. In another sense it *is* the same, since otherwise we wouldn't recognize it as *this*

[19]In ancient Greek it is often difficult to tell whether a category is being used abstractly (day, night, justice, strife) or is being personified (Day, Night, Justice, Strife, i.e. the goddesses Hēmera, Nux, Dikē, Eris); the two were not entirely separated in ancient thought. I've capitalized the words where personification seems to reinforce the meaning.

[20]The original text is not verse; I've formatted the text is in this way to try to reproduce the ambiguity of the Greek by weakening the connection between 'self same' and 'streams' and by strengthening the connection to those doing the stepping. Again, however, it is better to read the text without trying to parse it.

stream. Further, we must recognize that there is an essential unity in this opposition of flux and stability, since the water must flow, or it isn't a stream (e.g., it would be a stagnant puddle): its identity *is* a flux. Heraclitus takes the stream as the model for all stable objects: a balanced flow resulting from opposing forces in equilibrium.

It might seem that I am "overinterpreting" this fragment, but this is the way it was understood in antiquity:

(33) According to Heraclitus one cannot step twice into the same river, nor can one grasp any mortal substance in stable condition, but by the intensity and the rapidity of change it scatters and again gathers. Or rather, not again or later but at the same time it forms and dissolves, and approaches and departs. (Plutarch *Concerning the 'E' at Delphi* 392B) (tr. Kahn)

Heraclitus has hidden additional meaning in Fr. 32, since the phrase 'self same' ambiguously modifies both the streams and those doing the stepping. In the first case Heraclitus shows us that a person cannot step twice in the same *stream*, but in the second that the same *person* cannot step twice into the stream, since the person, like the stream, has changed. Evidence that this ambiguity is intentional is provided by:

(34) Into the same rivers, we step and do not step, we are and we are not. (tr. Kahn)

The claim here is *not* that we exist and we do not exist; the existential use of 'to be' had not yet evolved in Heraclitus' Greek (Kerferd, *SM*, pp. 94–96).[21] Rather, the meaning is that we are and are not a *particular* thing; we are the same, but also changed.

Oscillation

Another fragment illustrates a different way in which a unity may result from opposing forces:[22]

(35) They don't comprehend
how differing
with itself agrees:
back-turning *harmonía*

[21] For a comprehensive treatment see Kahn (*Be*).

[22] Again, this is not verse; the intent is to use the line break to weaken 'differing with itself' and to strengthen 'with itself agrees' (since the latter is inverted), thus reproducing the balanced ambiguity of the Greek.

as of bow and lyre.

The interpretation of this fragment depends crucially on the meaning of *harmonia* in Heraclitus' time. Its principal meaning is a joining or fitting together, as in carpentry. This is metaphorically extended into political usage to mean a pact or agreement. It is also used as a technical term in music to refer to a tuning, a musical scale or a melody. In all these we can see a common idea: a structure resulting from a fitting together or arrangement of complementary parts.[23] The whole has opposing components ("differing with itself") yet a coherent rational structure ('agrees' = *homologeei* < *homo* + *logos*).

Now consider how the harmonia of the bow and lyre are "back-turning," that is, turning in opposing directions or reversing. In a strung bow there is a balanced opposition of forces. Drawing the bow upsets the balance, and the arrow is propelled by the resulting motion in the opposite direction. The situation is much the same with the lyre, except that the effect we are interested in is exactly the back-turning motion — or oscillation — of the string. This is also suggested by 'differing' and 'agreeing': drawing apart and pulling together.[24]

Back-turning or Back-stretched?

There is disagreement about whether the bow and lyre are described as 'back-turning' (*palintropos*) or 'back-stretched' (*palintonos*). The adjective 'back-stretched' apparently makes more sense, but this in itself is a reason to reject that reading (since it is the *lectio facilior*[25]; see also Guthrie, *HGP*, Vol. I, pp. 439–440, n. 3). In fact, both words make sense, 'back-stretched' suggesting equilibrium and 'back-turning' suggesting oscillation (Hussey, *Presoc.*, pp. 44–45). Further, Kahn (*ATH*, pp. 195–196, 199–200) thinks that Heraclitus might be trying to suggest both ideas at once. Since 'back-stretched bow' is a common phrase in Homer, the occurrence of 'back-

[23] See Hussey (*Presoc.*, pp. 43–44), Guthrie (*HGP*, Vol. I, pp. 435) and Kahn (*ATH*, pp. 195–200) for more on *harmonia* in Heraclitus.

[24] The word translated 'differing' (*diapheromenon*) is literally 'drawing apart', and some readings of the text have 'draws together' (*xumpheretai*) instead of 'agrees'.

[25] A basic principle of textual criticism is *difficilior lectio potior*: the more difficult reading is preferable (Reynolds & Wilson, *S&S*, pp. 199–200). The reason is that as a text is copied and recopied there is a tendency to replace unusual words by common words and difficult ideas by simple ideas; thus the more difficult reading (the *lectio difficilior*) is often closer to the original than the easier one (the *lectio facilior*). We may say that the entropy of a text tends to increase; see Section 11.6.6.

turning' near 'bow' will *suggest* back-stretched while *saying* back-turning — an ingenious use of both resonance and linguistic density.[26]

Another Ambiguity

In Fr. 35 the phrase 'with itself' can modify either 'differing' or 'agrees' or both. On one hand, 'differing with itself' stresses a typical Heraclitean opposition underlying a unity: differing with itself is an agreement. On the other, differing 'agrees with itself' stresses that the opposition is a harmonia: a co-ordinated structure. Since Heraclitus could have easily made his statement unambiguous, he probably intended that both interpretations be activated simultaneously. Therefore the truest way to understand the fragment is to simply read it, without attempting to parse it.[27]

Fitting Opposition

The preceding interpretation of Fr. 35 is supported by Aristotle:

(36) Heraclitus says, "opposition is fitting, and from differences comes the most beautiful harmonia, and all things come to be through strife." (Aristotle *Nicomachean Ethics* viii.1. 1155b4)

Here 'fitting' (*sumpheron* = *xumpheron*) should be interpreted both literally as a convergence of the separated parts into a juncture (the original sense of harmonia), as well more metaphorically: opposition is appropriate, useful and helpful (extended meanings of *sumpheron*), and hence agreement (harmonia) may result from differing (*diapherontōn*). We should also recall, as would Heraclitus' readers, that Harmonia was the daughter of Ares, god of differing, strife and separation, and Aphrodite, goddess of attraction, amiability and the embrace (Otto, *HG*, pp. 95–103, 248).

Hidden Harmonia

As we have been warned before, however (Frs. 16, 17, 18; p. 364), we cannot expect the harmonia that governs the universe to be obvious:

(37) Unapparent harmonia is dominant over apparent.

We must look below the apparent structure of things to see their true, governing structure. (See also Kahn (*ATH*, pp. 202–204).)

Process More Basic Than State

In general we see that the "war" between opposites can result either in equilibrium, in which the forces are balanced, or in oscillation, a cyclic

[26] In other texts 'back-turning harmonia' refers to the rising and falling tones of a melody (Hussey, *Presoc.*, p. 45); composers still explain melody in terms of "back-turning" tension (Hindemith, *CMC*, pp. 87–89).

[27] The lyre and bow are attributes of Apollo — his instruments of pleasure and pain (Kahn, *ATH*, p. 200). Considering Fr. 19 (p. 364), it's perhaps not coincidental that Heraclitus chose these examples.

Harmonic Systems

Heraclitus recognized a fundamental property of harmonic systems that can now be explained mathematically. If in a system of opposed forces, the restoring force is proportional to the degree of imbalance (x), then the motion of the system is given by the differential equation:

$$\ddot{x} = -\kappa x.$$

This equation has the solution:

$$x_t = x_0 \sin(\kappa^{1/2} t + \alpha).$$

If the forces are initially in balance ($x_0 = 0$), then they remain in balance ($x_t = 0$); if they are initially unbalanced ($x_0 \neq 0$), then the system oscillates forever.

Heraclitus certainly recognized that most harmonic systems do not oscillate forever, due to friction and other energy dissipating processes. Often these processes are proportional to the rate of change, which is expressed in the differential equation:

$$\ddot{x} = -\kappa x - \lambda \dot{x}.$$

In this case there are two possibilities. If the frictional force is sufficiently small ($\lambda^2 < 4\kappa$) then the system is *underdamped*, and it oscillates indefinitely, but with gradually decreasing amplitude; this is the case with the lyre. If in contrast the frictional force is greater than the critical value ($\lambda^2 \geq 4\kappa$) then the system is *overdamped* and equilibrium is restored without oscillation.[28] This is what happens with the bow (at least if the string is not pulled so far that it vibrates). (For more, see for example Boas, *MMPS*, pp. 356–358.)

process in which first one extreme and then the opposite dominates. People give names to the phases of this oscillation, and don't recognize the unity of the underlying process:

(38) The same is there [*or:* is present in us]: living and dead, and the waking and the sleeping, and young and old; for these changed are those, and those changed again are these. (tr. after Kahn, Kirk, Raven & Schofield)

Universal Flux

The regularities of the universe — the stable states and objects as well as the cyclic processes — are a result of the war of opposites. In a summary of Heraclitus' philosophy this is expressed clearly:

> All (things) come to be through opposition, and the wholes flow like a river. (Diogenes Laertius, ix.8, 11–13 = DK 22 A 1])

Panta Rhei

Everything is in flux. Even apparently stable objects such as cliffs and mountains can be seen, if we interpret our perceptions with understanding (Fr. 11, p. 363), to undergo slow but continual change:[29]

(39) Somewhere Heraclitus says that all things move on and nothing abides, and, comparing existing things to the flowing of a river, he says you would not step twice in the same river. (Plato, *Cratylus* 402a)

(40) And some say, not that some existing things move and others not, but that all [things move] and always, though this escapes notice by our perception. (Aristotle *Physics* Θ.3. 253b9)

(41) It is out of movement and motion and mixture with one another that all these things become which we wrongly say "are" — wrongly because nothing ever is, but is always becoming... [Homer also] says: All (things) are born of flow and motion. (Plato *Theaetetus* 152d7–e8; tr. after Desjardins)

Universal Strife of Opposites

If it weren't for the *harmonia*, the coordinated opposition, of the forces, the universe would be structureless:

[29] There is no extant fragment in which Heraclitus states what is commonly attributed to him: "All things flow" (*panta rhei*); the source may be a paraphrase such as Diogenes' (p. 372), or Frs. 39 or 41, or indeed a lost part of Heraclitus' book.

(42) And Heraclitus rebukes The Poet [Homer], who said, “Would that strife might perish from among gods and men!” He did not see that he was praying for the destruction of the whole; for, if his prayer were heard, all things would pass away… (Aristotle *Eudemian Ethics* vii.1.1235a26 + Simplicius *Categories* 412, 26; tr. after Robinson)

Thus “the common,” which is the basis of true understanding (recall Fr. 25, p. 365), is to be found in recognition of this universal war:

(43) One needs to know that War is common and Justice is Strife, and that all things come to be in accord with Strife.

Since the strife between opposites is the common law of the universe, it is appropriate to identify it with Justice (or “The Indicated Way” — *Dikē*).

10.2.7 Nature

The Universal Process

The process underlying all oppositions, manifesting itself first one way then the other (and being named accordingly), Heraclitus calls, in a kind of pantheism, “god”:[30]

(44) The god: day night, winter summer, war peace, glut hunger; (it) alters as, when mingled with spices, (it is) named according to the pleasure (fragrance) of each.

The subject ‘it’ is not expressed; the thing mingled with the spices could be could be fire, olive oil or wine (all have their supporters), but the point is the same; our categories are based on superficial properties and ignore the underlying truth. This is stressed by a double ambiguity: lexical and syntactic (Kahn, *ATH*, p. 280). The word *hēdonē* normally means pleasure (as in ‘hedonist’), but may be used technically to refer to a flavor or perfume. Further, ‘of each’ is syntactically ambiguous, and may mean ‘of each spice’ or ‘of each person’. Thus the text supports two readings: the underlying reality is named according to the fragrance of each spice and according to the pleasure of each observer.

[30]The identification of the divine with the all-encompassing principle of the universe was common among the Presocratics; see for example Kahn (*ATH*, pp. 11–12, 19), Hussey (*Presoc.*, pp. 17–18), Kerferd (*SM*, p. 163) and Guthrie (*HGP*, Vol. I, pp. 382–383).

The sentence itself has a remarkable structure. The first clause comprises nine nouns with no explicit verb, including four pairs of cyclic oppositions; these all denote *states*. The second part comprises three verb phrases with no explicit subject; they all denote *processes* (Kahn, *ATH*, pp. 276–281). The sentence structure suggests that states and properties are essentially illusory and that the true nature of reality is a universal process. (See also Popper, *C&R*, pp. 144, 159–165.)

Fire the Universal Agent of Change

What is the nature of this universal process that underlies all of nature, but which is variously named after superficial differences, and variously named by different observers? The fragments answer quite clearly: it is *fire*.[31] In a summary of the doctrine:

(45) All things from fire arise and into fire resolve; all things come to be in accord with destiny, and through opposing currents existing things are made a harmonia [*hērmosthai*]. (Diogenes Laertius, ix.7)

On the surface this idea appears quite naive, and we may smile at the optimistic simplicity of Milesian physics, but a closer look reveals an unexpected depth.

We have seen that all the structures of the world, whether objects or states, are processes resulting from oppositions, either in equilibrium or in oscillation. "All are born of flow and motion." (Fr. 41) What is the agent of this motion?

Evidence of Heat Energy

To understand Heraclitus' theory we must put aside for a moment our knowledge of modern physics, try to view the world through ancient Greek eyes, and ask what they would see as the universal agent of change. The two kinds of change that were most under experimental control at that time were *cooking* and *metallurgy* (Hussey, *Presoc.*, p. 51), in both of which change is brought about through fire; according to an ancient gnome, *ignis mutat res* — fire changes things (Prigogine & Stengers, *OOC*, p. 103).

Further, when the ancients watched the changes in the sky, they saw the sun, the moon and the stars, all apparently manifestations of fire. Finally, observing the familiar changes in their coastal environment, they saw the morning sun raise mist (*aer*) from the water; the mist in turn was dissipated into the fiery upper sky (*aithēr*) by the heat of the day (Kahn, *ATH*, p. 17).

[31] This, incidentally, supports the interpretation of Fr. refBy36 that says it is fire to which the spices are added, and which give it its fragrance.

The temperature, turning back toward the cold, allowed the fog to condense and rain to fall back into the sea:

(46) As Heraclitus said, death of fire is birth for fog, and death of fog is birth for water. (Plutarch, *Concerning the 'E' at Delphi* 388D–E)

Similarly, the variation of heat governs the changes between liquid water and solid ice, as well as the silting up and erosion of the coast (Kahn, *ATH*, pp. 17–18, 144):[32]

(47) Fire's turnings: first sea, but of sea one half earth and the other half "burner" [lightning storm]. Earth is dispersed as sea and measured in the same logos [proportion] as before it became earth.

Heat and Life

Finally heat, especially the sun's heat, is obviously essential to life, and the higher life-forms reveal a warmth that suggests an inner fire. In modern terms, energy is the agent of change, but to the ancients heat energy was the only visible kind of energy, and fire is the obvious symbol of heat energy.[33]

The Universe a Fire

The foregoing discussion may have made more plausible the idea that the universe is a fire, or, as we might say, a thermodynamic process. But we have already seen (p. 368) that all processes are governed by a "back-turning harmonia," so the transfers of heat energy are lawful (or proportionate, *ana-logos*). Heraclitus states this clearly:

(48) The world-order — the same for all — no one of gods or men made, but it was ever and is and shall be: fire everliving, being kindled in measures and in measures being extinguished.[34]

[32]The meaning is obscure, partly due to doubt about the meaning of "burner" (*prēstēr*); it may refer to the water sucked into the sky by a waterspout accompanied by lightning. See Guthrie (*HGP*, Vol. I, p. 463, n. 3), Kahn (*ATH*, pp. 139–143), Burnet (*EGP*, pp. 148–149) and Hussey (*Presoc.*, p. 53). There is also doubt about the text; the second sentence may begin, "Sea pours out from earth ..."

[33]Light would not have been recognized as different from radiant heat energy. The potential energy of gravity is so pervasive that it probably was not viewed as an agent of change. Further, on the surface of the earth, it takes heat (typically in the form of fire or a living thing) to raise an object (even water, as mist) in the gravitational field. The "way up" is fueled by heat; the "way down" would be viewed as a natural restoration of equilibrium (as in the bow or lyre).

[34]The universal fire is described as 'everliving' where we might have expected 'everlasting'. Since the gods represented the permanent forces of the universe, immortality

That is, the process is governed by quantitative laws, defined in terms of heat exchanges:

(49) All things are an exchange for fire and fire for all things, as goods for gold and gold for goods.

Thus heat energy, or energy in general, is the universal medium of exchange for all transformations (a very modern view[35]). This is the logos Heraclitus promised us in Fr. 1; although he pointed the way, we now recognize it for ourselves:

(50) Listening not to me but to the logos, it is wise to agree that all things are one.

The universe is a unified, lawful process, of which the familiar objects, states and changes are superficial manifestations. Even the sun (a god) is bound by the logos of the universe:

(51) Sun will not overstep his measures, otherwise the Furies, assistants of Justice, will find him out.

Strife, the war of opposing forces, compels all processes to obey its quantitative laws; the Furies are the enforcers of the "lawful way" (justice) and the punishers of transgressions of the law. As lightning was sometimes called "the strife of Zeus" (*eris Dios*), we should not be surprised to find:

(52) [Heraclitus says,] "Thunderbolt steers all things" (regulates all things); by 'Thunderbolt' he means the eternal fire, ... and he calls it Need and Glut. (Hippolytus, *Refutatio* ix.10)

The thunderbolt can easily be imagined as the purest form of fire; since it's also the principal attribute of Zeus this fragment resonates with Fr. 44 (p. 373). Additionally it is identified with the principal opposition: deficiency and excess (Need and Glut); recall also Fr. 43.

Objects are Flames

was their defining characteristic. Thus this use of 'everliving' suggests again Heraclitus' pantheistic identification of the divine with the universal process (cf. Fr. 44, p. 373). We will see below (p. 377) that Heraclitus probably took life to be a kind of self-sustaining thermodynamic process — temporary in the case of mortals, perpetual in the case of the "divine" cosmic fire.

[35]Specifically, whenever a body moves in a force field, energy is exchanged between the body and the field.

We saw in Section 10.2.6 that all objects are like streams: temporary equilibrium states in the war of opposites. Now we understand that the underlying process is not one of flowing water, but of flowing heat energy. We may say that *objects are like flames*, since, like them, they are processes that are maintained by a balanced energy exchange.[36] Although this is not stated explicitly in the extant fragments, it seems to be the implication of the combined flux and fire doctrines, an interpretation defended vigorously in Popper (*C&R*, pp. 160–162) and Popper (*OS&E*, pp. 472–473). It was also Diogenes Laertius' interpretation, as we saw before (Fr. 45).

Near-Equilibrium Systems

The structure of some objects is a relatively simple harmonia that maintains their equilibrium state. For example, a rock represents a stable balance of forces (interatomic forces, to be specific) and will retain its identity for a long time, although its atoms are in continual oscillatory motion, and in the end the rock will crumble.

Nonequilibrium Systems

Other objects are more complex and require a constant energy input to prolong their existence; now they are called *nonequilibrium systems*. Heraclitus provides an important symbol of these:

(53) Even the *kukeon* separates unless stirred.

Kukeōn is was the sacred drink of Eleusis and was made by mixing barley meal and pennyroyal with wine, milk or water, often with additions of honey, salt or herbs (Donnegan, *Lex.*, s.v.; Kerényi, *El*, App. I). Such a drink requires constant stirring — constant energy input — to remain mixed. Like the stream (Fr. 32), the kukeon is what it is only by virtue of its motion (indeed the word *kukeōn* comes from *kukaō*, meaning to stir, mix or churn). It is the nature of such objects to be constituted of both flow of energy and motion of matter.

Life as a Flame

The flame-like nature of nonequilibrium systems is most apparent when we come to living things, for, like a flame, a living thing must be fed and needs air to survive; also they both give off waste in the form of heat energy and oxidation products (gaseous and solid). These observations may have led Heraclitus to conclude that the *anima*, or principle of life (see below), is like a flame. For example, compare Fr. 46 (p. 375) to:[37]

[36] They are regions of temporary stasis in the ever-increasing entropy guaranteed by the Second Law of Thermodynamics.

[37] Recall (p. 108) that for the Greeks a clear awareness was *dry* and impaired awareness was *wet*.

(54) For animae it is death to become water, and for water death to become earth. Out of earth water is born, and out of water anima.

This certainly suggests a close connection between anima and fire (Guthrie, *HGP*, Vol. I, p. 433), an interpretation supported by:

> By fire he does not mean flame: fire is the name he gives to the dry exhalation, of which the anima also consists.[38]

The "dry exhalation" is also suggestive of heat energy, perhaps with an emphasis on its transfer by convection.[39]

The Dry Anima

In support of the characterization of life as a flame, consider:[40]

(55) A radiance is the dry anima, wisest and best.

The sense seems to be that the wisest and best anima is a dry anima (a view supported by several other fragments), and that this is a radiance (*augē*, also flare, flash, sheen or gleam of sunlight), which supports the connection with fire. (Also, the sunlit sky was a traditional symbol for life (Kahn, *ATH*, p. 245–247).)

The Self-Increasing Anima

The natural tendency of fire is to spread, and so also for anima:

(56) The anima has a self-magnifying logos.

This fragment may refer to an obvious principle (logos) of life: when adequately "fueled" it grows and spreads; the word translated 'magnifying' (*auxōn*) means promoting, as well as increasing and strengthening, all of which suggest a kind of biological "will to power." (See also Section **??**.)

Life a Nonequilibrium Process

---

[38] Philoponus, commenting on 405a25 in Aristotle's *De Anima*; tr. after Guthrie (*HGP*, Vol. I, p. 432)).

[39] There is debate about whether Heraclitus thought the anima is (or is like) a fire or the "dry exhalation": see for example Guthrie (*HGP*, Vol. I, p. 466) for fire, and Kahn (*ATH*, pp. 239–240) against it. In either case the anima is closely connected with thermodynamic processes.

[40] See Kahn (*ATH*, p. 245–247) for a defense of the reading *augē xērē* in preference to *auē*. The meaning of the fragment is unclear, for the sentence comprises two nouns and three adjectives, all agreeing. Although 'dry' could modify 'radiance' or 'anima', the latter makes the most sense (Kahn, *ATH*, pp. 246–247). It is perhaps best treated as an appositional cluster connecting the ideas 'radiance', 'dry', 'anima' and 'wisest & best'; it is all resonance and no syntax.

Now it is a commonplace that life is a nonequilibrium thermodynamic process (see for example Prigogine & Stengers, *OOC*, pp. 175–176, and Jantsch, *SOU*, Chs. 2, 10), but Heraclitus could not have put it in these terms. He could however have reached essentially the same position by observing the similarity of flames and living things. That he could anticipate the idea of a self-sustaining nonequilibrium thermodynamic process is evidence of the power of critical speculation (Popper, *C&R*, Ch. 5).

Anima and Psyche

The word 'anima' requires some discussion, which illutrates the importance of connotation in translation. The Greek word *psuchē* ($\psi\upsilon\chi\acute{\eta}$) is often translated 'soul' but this is very misleading. Originally it seems to have meant the breath, especially as necessary for animating the body; in Homer, when a hero dies, his *psuche* escapes through his mouth. It is not especially associated with the understanding, the will, or the personality, all of which are usually referred to by other terms (*nous, thumos, phrenes* etc.). In early usage at least, *psuche* is mostly associated with the nonrational, biological aspects of animate life: initially with autonomic processes (breathing, blood circulation), later also motion, and finally sensation (see also Section **??**).

Thus *psuche*, certainly in Heraclitus' time, is very biological and very physical — even physiological. 'Soul' gives an entirely wrong idea, and 'spirit', which is etymologically good (cf. inspire, expire), has all the wrong connotations; even the transliteration *psyche* sounds excessively nonmaterial. Unfortunately the various phrases that might be used (life-breath, life principle, vital principle) are either too narrow or too abstract, and so would misrepresent Heraclitus' thought. These problems are avoided by keeping the word in Greek, but for many readers *psuche* has *no* connotations.

In the face of these difficulties I've chosen to translate *psuche* by the Latin word *anima*. First, it has a very similar manifold of uses, and has been used for centuries to translate *psuche*. Second, it has the correct resonance: it is the anima that animates the animal and without which it is inanimate. Although it's not an entirely satisfactory solution, I've found none better.[41] For more on *psuche* in the sixth and fifth centuries see Onians (*OET*, Pt. II, Chs. I, III), Peters (*GPT*, pp. 166–169), Kahn (*ATH*, pp. 126–127, 311, 237–243), Snell (*DMGPL*, pp. 8–22), and Dodds (*GI*, pp. 138–139).

No Complete Rational Account of the Anima

Heraclitus seems to have claimed that we will never have a complete

[41] Needless to say, the present use of *anima* has no direct relation to Jung's. For the Romans the *anima* was the vital principle and the *animus* was the principle of consciousness (Onians, *OET*, Pt. II, Ch. III).

rational account of the anima:

(57) The bounds of the anima you would not find out traveling every way, so deep a logos has it.

The bounds (*peirata*) are the limits and boundaries that make the logos (de)finite (cf. *apeiron*, p. 27). Much of Part III deals with the problems of finding a logos for the anima.

* * *

Fire the Judge of All

All things, living and nonliving, are like flames, temporary equilibria in the thermodynamic process of the universe. This process is lawful, however, and the law governs the evolution of the universe. Some flames will flare up, others will die down, and some will burn invariably, but the universal process selects some for survival and some for dissolution:

(58) The advancing fire will select and apprehend all.

Even the ineluctable forces of nature (the gods) are governed by the universal process, which grants perpetual existence to them, but not to mortals:

(59) War is father of all but also king of all, and reveals some to be gods but others men, and makes some slaves but others freemen.

The conflict of opposing forces is the source of all things, but also their governing principle.[42]

Nature Has No Values

Some things, animals, people and institutions survive, others pass away. People, with their "private thinking" (Fr. 22), or even with their shared but parochial customs (Fr. 25), take some events to be good and some evil, but from the standpoint of the cosmos they are neither; they are simply the inevitable evolution of the universal process:

(60) To god all things are fair and good and just, but men have taken some to be unjust and others just.

Justice *is* the universal fire: "War is common and Justice Strife" (Fr. 43).[43]

[42]By calling War the father and king of all in close proximity with a mention of gods and men, Heraclitus has established resonance with the Homeric formula for Zeus: 'father of men and gods' (Kahn, *ATH*, p. 208). In this way Heraclitus has reinforced the message that only the universal thermodynamic process is worthy of being called divine; recall Frs. 44 (p. 373), 48 (p. 375) and 52 (p. 376).

[43]Plato has his own description of "thermodynamic justice":

10.2.8 Conclusions

The intricate tapestry of Heraclitus' thought defies summarization; nevertheless we must try.

We have found that the familiar categories of thought will mislead rather than help, for they are based on superficial similarities and differences, and ignore the underlying unity of the universe. Although these categories allow the phenomena of the world to be divided and redivided, and like Pythagoras we may be very learned in the resulting logoi, we will ultimately lack understanding. True wisdom comes in a recognition of the basic process of the universe, which cannot be adequately described in words, although it can be indicated through gnomes and aphorisms.

If we try to put this insight in modern terms, we may say that the universe is a thermodynamic process, resulting from temperature differences (oppositions of hot and cold), and which in turn lead to other oppositions. Opposites may be in balance or in oscillation, but in either case the persistent structures, both spatial and temporal, are just the superficial manifestations of the heat-driven process. The quantitative laws of the universal process determine the fate of all such structures.

Nonequilibrium systems, like flames, maintain their structure through constant energy input. Of these, the premier examples are living systems; unlike *kukeon*, which must be stirred, organisms are self-structuring and self-nourishing. Understanding the anima in all its functions, including cognition, depends on abandoning our comfortable categories and recognizing the underlying process for what it is.

Rereading Heraclitus

I strongly recommend that you reread this section; remember the limitations of a linear presentation of a multidimensional manifold of ideas. Having read and pondered the later fragments, you will now find that many of the earlier ones make more sense; there are additional resonances that were not possible on first reading. Needless to say, even two readings will not exhaust the fragments of Heraclitus' book. You will not find out its bounds, even traveling every way, so deep is its logos.

> "One of them says it is the sun that is just (*dikaion*); for it alone by moving through (*dia-ion*) and burning (*kaon*) administers all beings... [Some other] says it is fire... Another says it is not fire itself but the very heat which is present in fire." (Plato *Cratylus* 413b–c; tr. Kahn, *ATH*, p. 275)

Notice the use of speculative etymology as a means of establishing resonance.

Correpondence of Heraclitean Fragment Numbers

M = MacLennan, **By** = Bywater, **DK** = Diels & Kranz 22B,
KRS = Kirk, Raven & Schofield

M	**By**	**DK**	**KRS**
1	2	1	194
2	52	61	199
3	51	9	
4		13	
5	53	37	
6	58	58	
7	69	60	200
8	50	59	
9	49	35	
10	13	55	197
11	4	107	198
12	3	34	
13	5	17	
14	16	40	255
15	17	129	256
16	18	108	
17	10	123	208
18	7	18	210
19	11	93	244
20	12	92	245
21	19	41	227
22	92	2	195
23	91a	113	
24	95	89	
25	91b	114	250
26	60	23	
27	104	110+111	201+
28	59	10	203
29	39	126	
30	35	57	

M	By	DK	KRS
31	83	84a	
32	42	12	214
33	81	49a	
34	40	91	214b+
35	45	51	209
36	46	8	
37	47	54	207
38	78	88	202
39		A 1	
40	62	80	211
41			215
42			216
43			
44	36	67	204
45			
46	25	76	
47	21	31	218
48	20	30	217
49	22	90	219
50	1	50	196
51	29	94	226
52	28+24	64+65	220+
53	84	125	
54	68	36	229
55	75	118	
56		115	
57	71	45	232
58	26	66	
59	44	53	212
60	61	102	206

10.3 The Sophists: Word Strife and Custom

> The Greek culture of the Sophists had developed out of all the Greek instincts; it belongs to the culture of the Periclean age as necessarily as Plato does *not*: it has its predecessors in Heraclitus, in Democritus, in the scientific types of the old philosophy; it finds expression in, e.g., the high culture of Thucydides. And— it has ultimately shown itself to be right: every advance in epistemological and moral knowledge has reinstated the Sophists— Our contemporary way of thinking is to a great extent Heraclitean, Democritean and Protagorean: it suffices to say it is Protagorean, because Protagoras represented a synthesis of Heraclitus and Democritus.
>
> — Nietzsche (*WP*, §428)

> Be neither saint nor sophist-led, but be a man.
>
> — Matthew Arnold, *Empedocles on Etna*, 1825, I.ii.136

10.3.1 Background

Meaning of 'Sophist'

Heraclitus showed the pitfalls of attempting to understand through the familiar categories the true structure of the universe; he argues that there is a deeper, hidden process that steers all things. We will see that sophism takes a similar view, but seeks its implications for human life and society. It is important to recognize, however, that 'sophism' does not refer to a school of philosophy, but to a profession. The term derives from *sophos* ($\sigma o \phi \acute{o} \varsigma$), which means 'skillful' or 'wise', and is related to words such as 'wisdom' (*sophia*) and recalls The Seven Wise Men (*Hoi Hepta Sophoi*). Thus we might call the sophists 'wisdomists' and say that they practice 'wisdomism'! The implication is that the sophists were (or professed to be) experts in wisdom and that sophism was the study of wisdom. That there could be such experts or such a study was a radical claim, as perhaps it still is. It will be worthwhile to consider some of the reasons.[44]

Aretē

[44]The word *sophos*, usually translated 'wise', originally referred to one who was clever or skilled in any art or craft; even a ditch-digger or hedger might be *sophos*. Later it came to mean wise, prudent or clever in practical matters, especially statesmanship; in the last

A crucial concept in this connection is denoted by the Greek word *aretē* (ἀρετή), often translated 'virtue', but perhaps closer in meaning to 'excellence'. Arete may be used to refer to "the strength of a lion, the fleetness of a rabbit, the sharp cutting edge of a pruning hook" (Wheelwright, *Pres.*, p. 237). Traditional Greek thought held human *aretē* to manifest itself in four cardinal excellences: self-mastery, courage, justice and wisdom.

It was also taken for granted that *arete* was an inborn trait, inherited from father by son, ultimately originating in a god. It was thus a matter of nature rather than nurture, a very convenient position for the aristocracy, who were called "The Beautiful'n'Good" (*Hoi Kaloi Kagathoi*). Against this we find the sophists claiming to provide (for a fee) what was assumed to be a gift of the gods.

Democratic Orientation

Sophism was not an isolated phenomenon; it was closely coupled with the fifth century (BCE) social and political developments known as the Greek Enlightenment. In particular, the progress of democracy was eroding the assumption that the right to rule was god-given and that arete was innate characteristic of the individual. For democracy to work the common people would have to learn the excellences, especially in public affairs, which had been taken to be the exclusive province of the aristocracy. The need for instruction in arete was filled by the sophists.

Reaction Against Sophists

It is hardly surprising that those with antidemocratic opinions denounced the sophists and accused them of fraud for claiming to teach what cannot be taught. More generally, conservatives, including many of the common people, saw the sophists as a threat to traditional Greek values. In particular, sophist skepticism about the traditional morality and religion invited accusations of atheism, and in the Enlightenment (!) we find a rash of prosecutions for "impiety." "The victims included most of the leaders of progressive thought at Athens, Anaxagoras, Diagoras, Socrates, Aspasia, Protagoras, and Euripides, though in his case it looks as though the prosecution was unsuccessful"

---

stage it meant wise and learned in general. The word *sophia*, wisdom, has much the same history, first meaning skill in arts and crafts (carpentry, music, medicine, etc.), and later learning, intelligence, sound judgement and wisdom, but always with a connotation of the practical. In the same way the derivative term *sophistēs* (sophist) originally meant an expert or master of some craft, and might be used of poets, musicians or cooks. Later it referred to wise, prudent, statesmen-like men, such as the Seven Wise Men, the presocratic philosophers, and — finally — the sophists. See Liddell, Scott & Jones (*LSJ*, s.vv. σοφός, σοφία, σοφιστής).

(Kerferd, *SM*, p. 21).[45] According to ancient sources (Diogenes Laertius, ix.52), copies of Protagoras' *On the Gods* were collected from all who owned them and were burned in the market-place at Athens; Protagoras himself was exiled and died enroute to Sicily.[46] Other sophists were punished by imprisonment, exile or execution.

> Dark-brow'd sophist, come not anear:
> All the place is holy ground.
>
> — Tennyson, *The Poet's Mind*, 1830

Socrates a Sophist?

Again, it is important to keep in mind that the sophists did not constitute a school of philosophy; they were teachers for hire in competition with one another. Dramatic representations of the sophists (e.g., by Plato and Aristophanes) show them promoting their own "products" and disparaging those of the competition. There is no doubt some truth in this, and it may help explain the vehemence of Socrates' attack on the sophists: he was one of them. Certainly he was concerned with many of the same problems as were they, and his methods had much in common with theirs. Indeed, he differs from them in only two major respects. First, he did not accept money for his teaching, which may reflect his antidemocratic inclinations, since he denigrates the sophists for being willing to teach arete to "all kinds of people," so long as they can pay (Kerferd, *SM*, pp. 25–26; Stone, *ToS*, pp. 40–41). The second difference is more to the point, since Socrates claims that arete ultimately derives from knowledge of the forms, the kind of theoretical knowing we explored in Parts I and II (recall Section 2.4). In this section we will see the sophists' very different account of arete, knowledge and education, and of their role in society.

Role in Later Philosophy

Little of the sophists' writing survives. Certainly public book-burnings did their part, but even more important was the vigor of the Socratic attack and the eloquence of Plato's presentation of it. The denunciations by Socrates, Plato, Aristotle,[47] and their successors have been so successful in

[45] Since Pericles and the sophists supported each other and were closely associated in the public mind, it's also possible that an attack on the sophists was seen as an indirect attack on Pericles.

[46] Some scholars (e.g. Burnet, *GPI*, p. 112, Hammond & Scullard, *OCD*, s.v. Protagoras) claim this story cannot be true; there is nevertheless "strong evidence" of a series of prosecutions (Kerferd, *SM*, p. 21).

[47] It should be noted that Plato and Aristotle were also no great friends of democracy.

their critique, that from their time until well into the last century the prevailing opinion has been that the sophists were not "true philosophers." The terms 'sophist' and 'sophism' were once terms of praise, but in our time still bear the negative connotations heaped on them by the Socratic offensive. If philosophy is just "footnotes to Plato" (p. 41) then the importance of the sophists may be seen in the number of Platonic dialogues devoted to refuting specific sophistic doctrines. The sophists may not have been "true philosophers," but they nevertheless set the agenda for philosophy for much of the next two and a half millennia. Now that the limitations of the Pythagorean/Socratic doctrine are known, the sophists may once again have their day.[48]

10.3.2 Antilogic

Rhetorical Self-Consciousness

Words, arguments, explanations — in short, *logoi* — were of increasing importance in fifth-century Greece; the reasons include the spread of written constitutions and laws, the development of explicitly formulated philosophical principles, and a heightened awareness of the power of literature. Accompanying these changes was an "awakening of what has been called rhetorical self-consciousness" (Kerferd, *SM*, p. 78), an awareness that logoi are not merely reflections of reality, but that they have a causal efficacy all their own.

Rhetoric and Eristic

One outcome of this awakening was the development of *rhetoric*, the art of persuading by any available means, from organization of argument to tone of voice or gesture. A particularly important kind of rhetoric was *eristic* (from *eris*, strife), the art of winning "contests of words." In this case the goal is

[48]As usual, the fragments are in Diels & Kranz (*Frag.*), but the translations in Freeman (*APSP*) omit the *Twofold Logoi* and *Anonymus Iamblichi*, and omit or summarize the fragments from many of the sophists. A useful collection of translated fragments, along with accompanying commentary, is Robinson (*IEGP*, Chs. 12, 13, App. B). Adequate selections from the fragments can be found in Wheelwright (*Pres.*, Ch. 8) and Nahm (*SEGP*, pp. 220–264); the latter includes selections from *Anonymus Iamblichi*. A Greek text of the *Twofold Logoi*, with facing translation and useful commentary, is Robinson (*CA*). Two dialogues of Plato, the *Protagoras* and the *Theaetetus*, have reasonably complete and relatively unbiased presentations of Protagoras' sophism. The most useful secondary source for me has been Kerferd (*SM*), although the organization of this section is based on the reconstruction in Untersteiner (*Soph.*, Chs. II–III). A detailed analysis of the sophist movement is Guthrie (*HGP*, Vol. III, Chs. I–XI), and a brief overview is Hussey (*Presoc.*, Ch. 6).

to persuade the audience that one logos is better than another, obviously an important skill in the assembly and law courts, where it is still widely practiced. In these arenas verbal contests were conducted "no holds barred," so it was necessary that the contestants be adept at all means by which one logos might prevail over another.

Antilogic

One of the most distinctive tools of eristic was *antilogic* ($\dot{\alpha}\nu\tau\iota\lambda o\gamma\iota\kappa\acute{\eta}$), "the opposition of one logos to another either by contrariety or contradiction" (Kerferd, *SM*, p. 63).[49] One way the "antilogician" can proceed is to oppose the adversary's logos with a contrary one that is at least as strong. Alternately, by a process of cross-examination (*elenchus*) the antilogician may be able to show that the adversary's logos leads to conclusions that contradict other positions held by the adversary, who is thereby forced into supporting both sides of an antilogical opposition. This is the familiar proof by contradiction.

Antilogic, Science and Scholarship

It must be stressed that there is nothing *illogical* about antilogic, which pits logical arguments against one another in a contest of strength. In fact, science and scholarship in general can be considered *systematic antilogic*, since they depend on a continual "contest of words" between scholars holding opposing views. The role in science of antilogical processes (specifically, criticism, falsification and refutation) is discussed further in Sections **??** and **??**.

Destructive Nature of Antilogic

Notice that antilogic is essentially a destructive process; it weakens our confidence in a logos by showing that it is inconsistent with other accepted logoi, or that it is no more plausible than other opposed logoi.[50] Although antilogic is well suited to the adversarial ends of eristic, many people doubted whether it was capable of revealing the truth of a matter; demagoguery in the assembly and courts no doubt provided ample grounds for such doubt.

Dialectic: Cooperative vs. Competitive

Socrates was an implacable critic of eristic and antilogic, in place of which he proposed *dialectic*. Exactly what he meant by this term is not clear

[49] I have adopted Kerferd's suggestive translation of *hē antilogikē* in preference to more awkward phrases, such as 'the art of opposed logoi' or 'the art of contradiction'.

[50] *Opposition* is cited by Prier (*AL*, p. 3) as one of the principal characteristics of archaic (i.e. pre-Aristotelian) logic. This is most commonly an opposition of contrasting symbols, which by a dialectical process resolves into a third symbol unifying the opposition; we have seen this repeatedly in Heraclitus, where it is the process, strife (*eris*), that "takes on the unifying and connecting power of an underlying third term" (Prier, *AL*, p. 64). We find this same pattern of thought in antilogic, in which the opposed logoi are more often arguments or explanations rather than symbols. The unification of the opposition can be found in the resulting state of doubt or perplexity (*aporia*).

(Kerferd, *SM*, Ch. 6), but a helpful clue can be found in the everyday meaning of the word: *conversation* or *discussion*. Thus it may be that dialectic suggests *cooperation*, whereas eristic suggests *competition*; we will see later that cooperation and competition are complementary processes for establishing some logoi in preference to others.

Dialectic Also Destructive

In fact, Socrates' methods are not as different from the other sophists' as he would have us believe. Many of the dialogues show him cross-examining his fellow citizens, and he is not above using antilogic to prove a point (although he may do so apologetically, e.g., *Theae.* 164C9–D1). In addition, more often than not, the result of the Socratic process is not some insight into the truth, but a state of perplexity or doubt (*aporia*) in which everyone agrees that they don't know what they are talking about. Thus dialectic, at least as practiced by Socrates, is frequently destructive, and indeed the destruction of his fellow citizens' opinions is often his apparent goal. As Socrates was fond of saying, he was wiser than other men only in *knowing* that he knew nothing; he tries his best to teach others this insight. Plato attempts to use dialectic more constructively, as a means of charting the formal structure of knowledge (Section 2.4.4).

Twofold Logoi

No matter what the goal, skill in opposing logoi (antilogic) was important in both eristic and dialectic. To teach their students this skill, sophists presented examples of *twofold logoi*, pairs of opposed logoi on some topic, and they no doubt required their students to compose their own twofold logoi. An anonymous work known as the *Twofold Logoi* (*Dissoi Logoi*) may be such an exercise, or it might be notes on lectures presenting twofold logoi (Robinson, *CA*). In this work we find numerous opposed logoi concerning what is good or bad, seemly or shameful, just or unjust, true or false, and whether or not wisdom and arete are teachable.[51]

Precursor to Liar Paradox

Not surprisingly, the *Twofold Logoi* contains an argument related to the Liar Paradox (p. 227); it is not known which came first.

It is also said that the false statement is different from the true

[51]The composition of twofold logoi did not stop in ancient Greece; I have already mentioned (p. 67) Pierre Abélard's *Yes and No* (*Sic et Non*), which was a collection of contradictory positions by the church fathers on 158 important theological issues. He studied "the art of disputation wherever it was flourishing," was widely sought after by students, and is often considered the founder of the University of Paris. Nevertheless — or perhaps because of these and other sophistical activities — his writings were burned and he was convicted of heresy by the Council of Sens (1141) and later by the pope. He died the same year.

> statement; as the name differs, so likewise does the reality. For if anyone were to ask those who say that the same statement is false and true which of the two their own statement is, if the reply were "false," it is clear that a true statement and a false statement *are* two different things, but if he were to reply "true" then this same statement is also *false*. (*Dissoi Logoi* 4.6; tr. Robinson, *CA*, p. 125)

Misology

The practice of composing twofold logoi would no doubt make one suspicious of *all* arguments — a state Socrates calls *misology*, the hatred of logoi (*Phaedo* 89C–91C). The ceaseless wrangling of the natural philosophers can have only contributed to this suspicion and may have led to the sentiment expressed in the opening words of Protagoras' *On the Gods*:

> Concerning the gods, I do not know whether they are or not. For many are the obstacles to knowledge: the obscurity of the subject and the brevity of human life. (Diogenes Laertius, ix.51 = DK 80 B 4; tr. Robinson)

(This was the book that led to Protagoras' exile and has the distinction of being the first to be burned by government order Hussey, *Presoc.*, p. 116.) It is also possible that *On the Gods* was just a part of a book called *Opposing Arguments* (*Antilogiae*), in which he applied antilogic to all the major issues of philosophy: the gods, being, the laws, politics, and the arts and sciences (Untersteiner, *Soph.*, pp. 10–15). The apparent futility of trying to settle these issues rationalistically may have led Protagoras to seek a more secure foundation for knowledge: the phenomena of sense perception; we will explore his doctrine in the next section.

Zeno, Gorgias and Being

Zeno's paradoxes (Section 2.3) provide a good example of antilogic. He strengthened the otherwise rather weak logos of Parmenides, that "all there is is one, altogether and changeless," by weakening the opposing logos, that there are many, changing things. He did this by arguing that both continuous and discrete pluralities have inherent contradictions, as does change. His antilogical accomplishment was challenged later by Gorgias (fl. 427 BCE), who opposed Parmenides' logos of Being with his own logos of Not-Being; in his book (*On Not-Being*) Gorgias argued

> first, that nothing is; second, that even if it is, it is incomprehensible by men; and third, that even if it is comprehensible, it is

certainly not expressible and cannot be communicated to another. (Sextus Empiricus *Adversus Mathematicos* vii.65 = DK 82 B 3; tr. after Robinson)

10.3.3 Homo Mensura

> *Socrates:* Does it seem to you, Hermogenes, that this is true of existing things: that what a thing is differs with the individual, as Protagoras maintains when he says that of all things the measure is man — so that as things *appear* to me, so they *are* to me, and as they *appear* to you so they *are* to you? Or does it seem to you that they have some fixed nature of their own?
>
> — Plato, *Cratylus* 385E

Protagoras: c.490–c.421 BCE

Since sophism was less a school than an outlook, it will be helpful to focus in this section on one sophist, Protagoras. He was apparently the first to call himself a sophist and the first to accept money for teaching. If the surviving fragments are any indication, he may also have been the best of the sophists; it's not unusual for great thinkers to be improvements on most of their successors. Another reason to focus on Protagoras is that a fairly complete reconstruction of his philosophy can be based on the extant fragments together with the extensive — and apparently quite accurate — account of his ideas in Plato's dialogues. Of course, the result is still an interpretation, with all that that implies (Section 10.2.3).

Democritus: c.494–c.404 BCE

Ancient biographies claimed that Protagoras was a student of Democritus, the originator of atomic theory. More recent scholarship has cast doubt on this possibility, but it still seems likely that they were acquainted, or were influenced by each other's ideas (Untersteiner, *Soph.*, p. 2; Hussey, *Presoc.*, pp. 111–113). Certainly Democritus' theory provides the easiest perspective from which to understand Protagoras' doctrine.

Physics of Perception

Democritus explained natural phenomena in terms of the motion and interaction of indivisible elementary particles (*atoma*), much as we do today. In particular, perception is the interaction of particles emitted by an object with the particles constituting the perceiver's anima. Thus, familiar sensations such as *cold, sweet* and *red* result from particular kinds of interactions of the *object* and the *subject*. We are accustomed to attributing such properties to the objects themselves. In fact, however,

> Sweet exists by convention, bitter by convention, color by convention; but in reality atoms [indivisible particles] and void alone exist. (Sextus Empiricus, *Adversus Mathematicos* vii.135 = DK 68 B 9; tr. Robinson)

Sweet is not a property *of* the object; it is a convention of language to use 'sweet' to refer to similar interactions between the particles of objects and subjects.

Relativity of Appearances

Democritus' theory of perception may have been what Protagoras had in mind when he began his book *On Truth* with the words:

> Of all matters the measure is man; first, of those being, that they are, but also of those not being, that they are not. (Sextus Empiricus *Adversus Mathematicos* vii.60 = DK 68 B 1)[52]

But acceptance of Democritus' theory is not a prerequisite for belief in the relativity of appearances; the evidence is ready to hand: the same wind feels cold to one person, but warm to another. Further we observe that appearances depend on the bodily state of the perceiver: the same food tastes sweet to a healthy person but bitter to the sick. Indeed, the same object may simultaneously appear cool and warm to the same observer: hold one hand under the hot tap, the other under the cold. Then plunge both into lukewarm water; to one hand it's cool, to the other warm.

Being and Not Being

These observations show that we cannot say that the water *is* warm or *is* cool. We may follow Democritus and say that the water is *neither* warm nor cool; in fact it's just mutually interacting elementary particles. Protagoras, however, is not concerned with speculation about the ultimate constituents of nature; his concern is to teach his student "how to exercise good judgement in ordering both his own affairs and those of the city, and how to be a man of influence in public affairs, both in speech and in action" (Plato, *Prot.* 318E5–319A2). Protagoras' practical orientation leads him to the conclusion that the water is *both* warm and cool; it is warm for me and cool for you, or even for my other hand. Since the water simultaneously *is* warm and *is not* warm, we may say elliptically that it *is* and *is not.*[53] The water's *being* warm

[52] This fragment is repeated in several texts; see Sextus Empiricus, *Outlines of Pyrrhonism* 1.216, Plato, *Theaetetus* 152a and Diogenes Laertius, ix.51.

[53] Thus we are interpreting absolute uses of 'to be' predicatively rather than existentially. That is, when I say "*X* is," this is short for "*X* is so and so," not for "*X* exists." This seems to have been the meaning of the absolute construction in the fifth century (Kerferd, *SM*, pp. 94–95; Kahn, *Be*). For some modest doubts, see Rankin (OEA, pp. 27–28).

depends on the individual perceiver, and even the locus of perception (e.g., which hand). Ultimately, the only standard of *being* is the continuing flux of appearances, the *phenomena*. This seems to be what Protagoras means when he says (p. 392) that the perceiver is the criterion of those things *being* (so and so), that they are (so and so), and of those *not being* (so and so), that they *are not* (so and so).

Phenomena

A *phenomenon* is literally something that appears to the senses or imagination, something that is mentally apparent. It derives from the Greek *to phainomenon* (τὸ φαινόμενον), which means something that comes to light, appears, or seems to be; something manifest or evident to the senses, or that appears in sense experience or the imagination; something that is observed or comes about, and so forth (Liddell, Scott & Jones, *LSJ*, s.v. φαίνω, B). I will have much more to say about phenomena in Section 11.3.

All Appearances Are True

The same considerations lead Protagoras to a radically subjectivist theory of truth. Surely my claim that the water is hot is just as true as your's that it's not, once it is recognized that I must mean that it's hot *for me* and you must mean that it's not hot *for you*. And that is all we could mean, for we know that in fact the water is neither; it's just particles in motion. Democritus' theory supports this interpretation, for if 'hot' names anything, it names a certain class of interactions of the elementary particles of the perceiver and the perceived. If I perceive the water as hot, then of necessity that interaction is taking place, and so my claim is true. If you perceive the water as not hot, then that interaction is not taking place, and so your claim is also true.[54]

Contradictoriness of Being

It would appear that Protagoras is asserting the contradictoriness of the universe: everything *is* and *is not*. Anything that is true may be simultaneously false, if it so appears. What he has actually done however is to relativize being and truth to moment by moment perceptual interaction. If I specify that the water is hot *to me, at a specified time, on my left hand*, etc. etc., then there is no opportunity for contradiction, but as soon as we generalize from the particular phenomenon, the possibility — indeed the virtual necessity — of contradiction is with us. This is a most extreme kind of

[54] We saw before that the *phenomena* are literally what appears, or comes to light. The Greek word for 'true', *alēthēs*, means 'unconcealed', and is a negative derivative of a verb (*lēthō*) meaning 'to escape notice or be forgotten' (Liddell, Scott & Jones, *LSJ*, s.vv. ἀληθής, λήθω, λανθάνω). Thus in ancient Greek it is nearly tautological that the phenomena are true, since that is just saying that *what appears does not escape notice*, or *what comes to light is unconcealed*. (See also Heidegger (*EGT*, p. 103, fn.).)

subjectivism and would seem to lead to a very impoverished epistemology, but we will see later (Section 10.3.5) how Protagoras is able to recover a kind of objectivity.

All Beliefs Are True

Before I come to this, however, it's necessary to mention one more implication of Protagoras' extreme subjectivism, which follows from his claim that "the anima is nothing more than the sensations" (Diog. Laert., ix.51). In this he seems to be identifying the anima with all the phenomena that appear to it, or as we might say, with everything of which it's conscious. (Recall that Greek *to phainomenon* can refer to something that appears to the imagination as well as to the senses.) Since any phenomenon is a manifestation of the state of the brain, it is in that sense true (it corresponds to the actual motion of the elementary particles, because it is caused by them). And so, in the words of the Platonic Socrates:

> He says, doesn't he, that what is *believed* by each person *is so* for him who believes it. (Plato, *Theaetetus* 170A3–4; tr. Robinson)

Conversely one cannot believe what is not true! Such an epistemological victory is Pyrrhic indeed, since it comes at the cost of abandoning all objectivity.

Paradoxes of Subjectivism

Protagoras' doctrine has born a pack of paradoxes: One cannot believe what's not true, and everything that's believed is true. But it also follows that of everything there may be (equally true!) contradictory statements. This is because each perceiver is the standard of what is and what is not, and hence of what is true or false.

10.3.4 The Better Logos

The True vs. The Better

Socrates asks[55] Protagoras the obvious question: If what each person believes is true, then why should we pay the sophists large fees to teach us? Protagoras answers that if Socrates wants to understand then he must put aside his usual logic-chopping and try to understand Protagoras' intent. He goes on to explain that although all appearances are true, some may be "better" than others. The job of the sophist is to make the phenomena better for us, to make the better *be* for us. This was traditionally the job of the wise men. The Platonic Protagoras summarizes their task as follows:

[55] The source for the following is the so-called Defense of Protagoras (Plato, *Theaetetus* 166A–168C).

> These appearances some, through inexperience, call 'true'; but I say that some are better than others, but not 'truer.' ... But the wise man causes the good things instead of the bad to appear and to be for them in each case. (Plato, *Theaetetus* 167B1–C7; tr. Robinson)

Sophism as Therapy

Sophism is essentially therepeutic. For example, food may taste bitter to a sick person, but sweet to one who's well. Both perceptions are true, but it's better (more desirable) for the food to taste sweet. The doctor does not change the food, but through drugs cures the patient's body, and thus changes the food from *being* bitter to *being* sweet. The sophists' cures are much the same, except instead of drugs they use words (logoi). By administering the proper logoi, the sophist can change *being* for the patient, replacing worse phenomena (appearances) with better.

A Sophist Psychotherapy Clinic

Plutarch reports that the sophist Antiphon developed a kind of psychotherapy and practiced it in a clinic:

> He constructed a technique of curing distress of mind analogous to the doctor's therapy for sick people, and having set up a room for the purpose next to the market-place at Corinth, he put out a notice saying that he was able to treat distress of mind by means of logoi. He would inquire the causes from his patients, and then consoled them. But he decided this technique was beneath him, and turned to rhetoric instead. (DK 87 A 6; tr. after Hussey)

Proper Use of Rhetoric and Eristic

In effecting a cure, the sophist may make use of all the devices of rhetoric and eristic. The sophist and diplomat Gorgias, who was known in antiquity (f. 427 BCE) for his rhetorical skills, wrote:

> The power of speech over the disposition of the anima is comparable with the effect of drugs on the disposition of the body. As drugs can expel certain humors from the body and thereby make an end either of sickness or of life, so likewise various logoi can produce grief, pleasure, or fear, which act like drugs when they give rise to bad persuasions in the anima.[56]

Although the sophist's skills may be used for good or bad ends, the wise always use them for the good. In this regard verbal prowess is no different

[56]Gorgias, *The Encomium of Helen* 14; tr. after Wheelwright). For more on the power of rhetoric, see Plato, *Gorgias* 456A–457C.

from physical prowess: those who use their rhetorical skill for wicked ends are no better than verbal thugs:

> Whenever anyone has striven for one of these things and has succeeded in accomplishing and acquiring it, whether it be eloquence, wisdom or strength, he must use it for good and lawful ends. And if anyone shall use the good in his possession for unjust and unlawful ends, this is the worst sort of thing, and it is better for it not to be present at all. And just as he who has any of these things becomes perfectly good by using them for good ends, so he becomes correspondingly utterly bad by using them for wicked ends.[57]

Such people are the reason that rhetoric, sophistry and oratory have acquired their negative connotations. The true sophist tries to change *being* for the better.

The Better and The Stronger

But what is "the better"? It can be argued that the winning argument has shown itself to be stronger, and hence better, by virtue of winning. (The Greek *kreittōn* can mean both 'better' and 'stronger'.) This would be the view of the practicers of eristic, but Protagoras had a deeper understanding, which we explore next.

10.3.5 Nature and Custom

> He is a barbarian, and thinks that the customs of his tribe and island are the laws of nature.
>
> — George Bernard Shaw, *Caesar and Cleopatra*, 1898, Act II

> Custom that is before all law, Nature that is above all art.
>
> — Samuel Daniel (1562–1619), *A Defence of Rhyme*

The Rule of Nature

"The better logos" is explained in the Myth of Protagoras (Plato, *Prot.* 320C–322D = DK 80 C 1). The first phase of human evolution was governed

[57] Anonymus Iamblichi, 3.1, 2; tr. Nahm (*SEGP*, p. 256). This is a work by an unknown sophist, who is therefore called to as 'Anonymus Iamblichi'.

by *nature* (*phusis*), which he symbolizes by the Titanic gods. First the Titan Epimetheus gave humans and other animals their various physical capabilities (teeth, claws, wings, speed, etc.). Humanity, of course, did not fare very well in this distribution, so the Titan Prometheus attempted to rectify the situation by allotting to different people different talents in the arts and crafts.

Instability of Communities

It was difficult for individual people to survive the hardships of nature, so they tried to form communities so that they might take advantage of each other's skills and abilities. Unfortunately these communities always disintegrated because their members lacked a sense of right and respect for one another.

The Rule of Custom

Then human evolution entered its second phase, which is governed by *custom* (*nomos*) and symbolized by the Olympian gods. Zeus orders Hermes to distribute to all people a sense of right (*dikē*) and respect for others (*aidōs*). All or most people must have this innate sense, or stable communities will be impossible.[58] Anyone without the capacity for these basic virtues will die like a public pest.

Subordination of Nature to Custom

Other sophists explained some of the mechanisms by which communities may subordinate nature to custom. Callicles (Plato, *Gorgias* 482C) describes how the many, individually weak, constrain the strong by a combination of education and laws. He thinks that a sufficiently strong person would "burst his bonds" and reassert the dominance of nature over custom, but another sophist ("Anonymus Iamblichi" 6) says that no one is that strong; even one "tougher than steel in body and anima" would be restrained by the skill and power of the masses:

> Custom, that unwritten law,
> By which the people keep even kings in awe.
>
> — Charles Davenant (1656–1714), *Circe*, II.iii

Stabilizing Role of Religion

The sophists seem to have claimed that religion was another mechanism by which communities are stabilized. Sextus Empiricus (*Adv. Math.* ix.54 = DK 88 B 2) attributes this theory to Critias, who was not a sophist, but was one of the Thirty Tyrants and had at one time been a student of Socrates. The argument is that the laws restrain people from violent deeds

[58] Incidently, Protagoras uses this general distribution of a sense of right and respect as a justification for democracy.

only when they are likely to be caught and punished. Therefore, "some shrewd and wise man invented fear of the gods for mortals, so that there might be some deterrent to the wicked even if they did or said or thought something in secret." Of course we need not take seriously this postulation of an "inventor" of religion, any more so than that of a dispenser of respect and a sense of right. The point is that widespread belief in the gods — *regardless of the truth of the matter* — may serve to stabilize the community, and by customary belief restrain natural inclination. In sophist terms, although belief that the gods exist is as true as belief that they don't, belief in the gods may be "better" because life in the communal state is preferable to that in the natural state. It may also be "stronger" in that the community will not otherwise survive.

Necessity of Impiety Prosecutions

The irony is that this very explanation of religion creates skepticism about the gods, since it brings their existence into the realm of the conventional; the community's gods are as much a matter of choice as its laws. Thus an implication of the sophist theory is that the community will likely protect its customs by destroying the source of skepticism: the sophists. We may hope that some of the sophists prosecuted for "impiety" had at least the satisfaction of knowing why their community *had* to prosecute them.

Why Socrates Accepted His Punishment

Perhaps this is why Socrates accepted his fate with such equanimity; in the *Crito* (50a–54d) he imagines how the Laws and Constitution of Athens would rebuke him if he escaped his punishment:

> Can you deny that by this act which you are contemplating you intend, so far as you have the power, to destroy us, the laws, and the whole state as well? Do you imagine that a city can continue to exist and not be turned upside down, if the legal judgements which are pronounced in it have no force but are nullified and destroyed by private persons? (50a–b; tr. Tredennick)

Inculcation of Customs

The community impresses its customs on the child at a young age; every action of the child is reinforced positively or negatively, and its behavior is straightened "like a warped and twisted plank" (Plato, *Prot.* 325D6–8). This is followed by more formal education in which the child learns the community's myths, as well as practices such as music and athletics, "so that they become more civilized and more graceful and more in tune with themselves and more apt for speech and action" (*Prot.* 326B4–7). In adulthood the laws are a stabilizing mechanism, feed-forward in that they guide behavior, feed-backward in that they limit the effect of transgressions.

Nature and Nurture

The result is that communal (*politikē*) arete is a matter of both nature and nurture. Although the native sense of right and respect was given by Zeus, the customs are acquired through education. In this the sophists claimed that they had a vital role, for they had made themselves experts in teaching arete. As Protagoras said, "Learning requires both nature (*phusis*) and self-discipline. It has to begin when one is young. It does not take root in the anima unless it goes deep" (DK 80 B 3 + DK 80 B 11; tr. after Wheelwright). The potential must be inborn, but education and practice are required to actualize it:

> Whatever anyone wishes to bring to the highest perfection, whether it be wisdom, courage, eloquence, the whole of arete or some part of it, can be accomplished from the following: In the first place, there must be the natural capacity, which is a gift of fortune, but in addition a man must become a lover of the things that are beautiful and good, industrious and a precocious learner, passing much time with them. And if even one of these is absent, it is not possible to bring anything to the highest pitch; but when a man has all of these, whatever he may cultivate, in this he becomes unsurpassable. (Anonymus Iamblichi, 1; tr. after Nahm, *SEGP*, p. 255)

Sophist Educational Practice

Socrates believed that arete is a matter of having the relevant theoretical knowledge, specifically, an understanding of the logical structure of the form of the Good. For the sophists, in contrast, acquisition of arete requires natural aptitude, formal instruction and practice. Thus their educational methods augmented formal lectures with small discussion groups as well as question-and-answer sessions (cf. the so-called Socratic Method). Students were also given antilogical examples (such as the *Twofold Logoi*) and were expected to gain experience through their own rhetorical compositions. Students apparently lived with their teacher, perhaps in a master/apprentice relationship, so that they might emulate the character and intellectual style of the sophist, as well as benefit from the mutual stimulation of the other students. The excitement pervading sophist schools is depicted dramatically in Plato's *Protagoras*. A recurring theme in the following chapters will be the necessity of augmenting verbal instruction, as a means of learning *that*, with emulation, practice and reinforcement, as means of learning *how*. See Kerferd (*SM*, pp. 30–34, 134–138) for more on sophist educational techniques.

Custom by Nature Stronger

On the surface it appears that nature is opposed to custom, but a deeper understanding shows that custom is in fact better/stronger than nature: custom persists because it is — by nature — stronger. Although custom restrains nature, it is ultimately nature that gives it the power to do this. Protagoras says,

> For I claim that whatever seems right and honorable to a state is really right and honorable to it, so long as it believes it to be so; but the wise man causes the good, instead of that which is evil to them in each instance, to be and seem right and honorable. (Plato, *Theaetetus* 167C4–7; tr. after Fowler)

But this does not mean that these equally true beliefs are equally advantageous:

> But in establishing what is advantageous or disadvantageous to the state, there if anywhere, [Protagoras] will agree that ... the opinion of one state differs from that of others as regards the truth, and he would by no means dare to affirm that whatsoever a state establishes in the belief that it will be advantageous to itself will quite certainly be so. (Plato, *Theaetetus* 172A6–B3)

Thus, although what the community *believes* to be true will *be* true (for it), it will be of the *nature* of things that some such beliefs will allow the community to thrive and others will not. The standard of truth is in the short term social but in the long term pragmatic.

Certainty and Objectivity Found in the Common

Let's summarize. The ideal of knowledge has long been *certainty and objectivity*: certainty because it removes the mental discomfort of doubt, objectivity because it promises a standard of truth and an end to disagreement. Rationalism is a perennial temptation since it offers both, but the *Twofold Logoi* makes us doubt its ability to actually deliver them. Protagoras' "man is the measure" provides a different source of certainty, but it comes at the cost of the most extreme subjectivism. A measure of objectivity can be found in the common, the customary, the shared background. It's true because it's believed; it's objective because it's public (common, shared).[59]

[59] This grounding of objectivity in the common should be compared with Heraclitus' Fr. 25 (p. 365).

10.3.6 Conclusions

Cooperative and Competitive Dynamics

In the establishment of custom we can see a cooperative and competitive dynamic at work. In the first case we see this in the Myth of Protagoras, for it was the cooperation of people in communities stabilized by custom that allowed them to prevail over communities not so stabilized, as well as over the other hazards of nature. Custom itself evolves by a cooperative/competitive processes. Interpretations, beliefs and other logoi are set beside one another and by an antilogical process compete for people's allegiance. Logoi do this in part by spreading, reinforcing each other, and self-organizing into internally consistent world-views. These cooperative/competitive processes constitute adaptive mechanisms, the effectiveness of which decide the survival of these world-views.

Sophist Humanism

Sophism can be considered *humanistic* in several different respects. First it focused on *human* problems — how to conduct one's life, how to be successful, how to be a good and productive citizen – rather than on (at that time) irrelevant speculation about nature. Second, it based its understanding on *human* means of knowing — the phenomena as they appear to us — it eschewed rationalism and (at that time) unresolvable scientific debate. Finally, it made *human* institutions, the community and its customs, the basis for certainty and objectivity, to the extent that these are humanly achievable.

> The sophists agree in an anti-idealistic concreteness which does not tread the ways of scepticism but rather those of a realism and a phenominalism which do not confine reality within a dogmatic scheme but allow it to rage in all its contradictions, in all its tragic intensity, in all the impartiality imposed by an intelligibility which will revive the joy in truth. (Untersteiner, *Soph.*, p. xvi)

10.4 Aristotle: The Mental Faculties and Practical Knowledge

The remainder of this chapter exists only as a detailed outline.[60]

We have discussed Aristotle as the founder of formal logic (Sec. 2.5), and we return to him here for his biologically-grounded theories of cognition and for his account of practical wisdom, both of which are relevant to alternative theories of knowledge.

Reading:

1. Aristotle, *De Anima* 411a26–b30, 412a3–415a13, 418a7–419a15, 424a17–b3, 429a10–430a25 (Ross §§63-67). [*psychê*, mental faculties, perception, reason]
2. Aristotle, *De Sensu* 449a8–20 (Ross §65) [perception]
3. Aristotle, *De Memoria* 451a18–452a16 (Ross §68) [association of ideas]
4. Aristotle, *Nic. Eth.* Bk. VI. [kinds of knowing, practical wisdom]

10.4.1 Biological Basis of Mental Faculties

a. Nature of the Psychê

Meaning of psychê

Aristotle viewed living things as hierarchical arrangements of structures (systems, in modern terminology), each with its own characteristic activities and ends (goals). Aristotle used the term ψυχή (*psychê*) to refer to such a structure. So *psychê* is the *form* of a living object, and *psychology* is the study of such forms. (which is different from modern meanings). *Psychê* is usually translated 'soul', but that's not very accurate. In Ancient Greek ψυχή = reason, mind, intelligence, life, spirit, soul, heart, understanding (Liddell, Scott & Jones, *LSJ*, s.v.) Aristotle's definition: "*Psychê* is the first grade of actuality of a natural body having life potentially in it." The body is *organ*ized and its parts are *organs* (< Gk. ὄργανον = *organon* = instrument, tool). To ask whether body and soul (soma and psychê) are one is as meaningless as to ask whether the wax and its shape are one.

[60] See the Preface for an explanation.

b. Faculties of the Psychê

The powers of the *psychê* can be divided into several categories. Not all are possessed by all living things, which illustrates the danger of translating *psychê* as "soul." E.g., Aristotle says plants have a *psychê* since they have the power of self-nutrition. In addition to nutrition, animals have the faculty of perception, and besides nutrition and perception, humans have the faculty of reason.

i. Nutrition. The *nutritive psychê* controls the transformation of food into body parts in such a way that it fits the plan of the body. Therefore, in nutrition the matter of the food is taken into the body and molded by the form of the body. That is, the body takes in *matter without form.* This is essentially our modern view nutrition.

ii. Perception. The *perceptive psychê* is found in animal life. Here the sensible form of the object, rather than its matter, enters the body. The body takes in *form without matter.* A piece of wax has a potential for many shapes; when a seal is impressed on it one of those shapes is actualized. Similarly, the eye has a range of potentialities, which may be actualized by sense objects. The object itself, *qua perceivable object,* is also actualized, into a *perceived object.*

iii. Reason

Active and passive processes

Thinking is akin to perception in that the mind identifies and discriminates something that *is* (but sense depends on the body, whereas thinking does not). As in sensation, there are both active and passive processes involved. To accomplish this identification the mind, like the sense organs, must be able to take on the form of its objects. But in this case it is *intelligible form* rather than *sensible form.* Taking on the intelligible form is the passive process, which is error-free. The combination of concepts in judgement is the active process, and it is here that error can arise. This seems to imply that the recognition of things in conceptual terms is infallible, that is, that there can be no mistakes in identification, which seems counter to experience. Compare sight — it's passive; beyond opening our eyes, we don't have to *do* anything to see. Apprehending the intelligible form could be the same. We can't *fail* but note that Socrates is a man.

Source of knowledge

The exercise of higher faculties always depends on that of lower. Therefore the exercise of the intellect depends on prior sense perception. But this is not necessarily saying that all our ideas ultimately come from sense perception, which is (pure) *empiricism.* Rather, the higher activities depend on the working of the lower; e.g., we can't think without images.

Tabula rasa

But, since virtually anything may be the object of thinking, the mind can have no form of its own (for that form would color or limit our thinking). Therefore the mind potentially has *any* form. "Mind is in a sense potentially whatever is thinkable, though actually it is nothing until it has thought." Aristotle says the mind is like a blank tablet (*tabula rasa*). Here we have a key issue for the possibility and limits of knowledge: Is the mind a blank slate capable of taking on any ideas? Or does it come inscribed with ideas? Or does it have such a form that only certain ideas can be inscribed?

iv. Desire. Aristotle is generally content with the three faculties of nutrition, perception, and reason, but occasionally he cites a need for a faculty of desire (*orektikon*) as the ultimate initiator of motion. More specifically, purposive motion results from *practical reason* (see Sec. 10.4.2 below) applied to a desire. Of course, in the case of animals, this reasoning is imagistic rather than discursive, as is also the case often with humans.

c. Memory and Recollection

Memory

Aristotle distinguishes "memory" (Grk. *mnêmê*, Lat. *memoria*) from "recollection" (Grk. *anamnêsis*, Lat. *reminiscentia*). Memory is a function of the psychê's faculty of *phantasia* (Lat. *imaginatio*, "imagination"), which imprints the forms from sense perceptions (Grk. *phatasmata*, Lat. *imagines*, "images") onto the memory, like a ring's impressions in sealing wax or an inscription in a wax tablet. Although this impression is caused by the perception, it may be more or less faithful to the original. Aristotle notes that some things are remembered after only one exposure; others may be weakly impressed after many exposures.

Recollection

Recollection is the process of retrieving a memory, by starting with some mental image, and following the associations from remembered image to remembered image, searching through the image space until the sought image is recovered (see "Laws of Association" below). This is essentially a reconstructive process implemented by the imagination (*phantasia*), which is also responsible for perception, illusion, visualization, dreaming, and daydreaming. Indeed it is fundamental to all thinking (see "Visual thinking" below).

Platonic anamnêsis

Anamnêsis (recollection) is the term that Plato uses for the recovery of innate ideas (as in the *Meno*). Aristotle's theory of recollection would seem to include this, except that he rejects the notion of innate ideas (see "Tabula rasa," Sec. 10.4.1 above). A Jungian version of innate ideas is discussed in Sec. 11.4.3.

[Richard Sorabji, *Aristotle on Memory*, 2nd ed., Chicago: University of Chicago Press, 2006.]

d. Laws of Association

Aristotle formulated three laws of association (which are the basis of later theories of association):[61]

i. Law of similarity (or resemblance)
ii. Law of contrast (or opposition)
iii. Law of contiguity (in time) (or law of habit)
Aristotle is sometimes also credited with:
iv. Law of frequency

e. Visual Thinking.

Aristotle says, "The mind never think without an image," which looks forward to the importance of imagistic thinking as an alternative to discursive thought. [More precisely: "there can be no thinking (*noein*) at all without images (*phantasmata*) (*De Memoria* 450a1)]

f. The Art of Memory.

A review of the art of memory (discussed in Sec. 3.3.1, p. 77) as an example of visual thinking.

10.4.2 Kinds of Knowing

a. Pure Science

b. Applied Science

c. Practical Wisdom

i. General. Aristotle distinguishes theoretical wisdom (from *theôria* = contemplation, observation) and practical wisdom (from *praxis* = action,

[61] Aristotle, *De Memoria et Reminiscentia* (Ross tr., p. 111). They are anticipated in Plato's *Phaedo*.

doing). Theoretical reason focuses on the unchangeable and the true, whereas practical reason focuses on the changeable and choice.

ii. Concreteness of Practical Syllogism.

Practical syllogism

An example of a practical syllogism is:

	I should be healthy;
	if I exercise then I will be healthy;
therefore,	I should exercise.

In a practical syllogism, the major premiss ("I should be healthy") names the agent's goal or some good, and the minor premiss ("if I exercise then I will be healthy") is a means to that end or other necessary preconditions for it. The conclusion is the action to take (or subgoal to accomplish). It is presupposed that the major premiss is an attainable goal, and the minor premiss is a true proposition (often arising from the *nous* or intuitive faculty). In both theoretical and practical syllogisms the major premiss is more general and the minor premiss is more specific.[62] The conclusion of a theoretical syllogism is a proposition (what to believe), but the conclusion of a practical syllogism is an action or choice (what to do). Therefore the practical syllogism is often more relevant than the theoretical syllogism to AI and robotics, since the latter are often directed toward concrete action.

Parts of practical wisdom

Practical wisdom then has four parts:

1. a *general* conception of the good;
2. an understanding of how to achieve the good in a *particular* situation;
3. correct *deliberation* from the premisses to the conclusion;
4. the *ability to act* on the conclusion.

d. Intelligence

e. Theoretical Wisdom

[62] See Sec. 2.5 (p. 48) for examples of theoretical syllogisms.

Chapter 11

Epilogue

Unfortunately, the remainder of Volume II of *Word and Flux* will never be written. Arguably, it does not need to be written, because many of the ideas it intended to present are more widely known now. Nevertheless, I believe there is some value in providing a detailed outline of the second volume, if for no other reason than it provides a reading list and study guide. Each section of this chapter represents a chapter of Volume II, comprising the remainder of Part III, which is focused more on philosophy, and Part IV, which is focused on connectionism and neural information processing.

11.1 Mental Imagery

11.1.1 Hellenistic Psychology

Epicurean physicalist explanations of the mind.

11.1.2 Ordinary Language Philosophy

1. Wittgenstein: Family Resemblance and Language Games

a. Background

b. Nature of Language

i. Picture Theory

ii. Language Games

iii. Family Resemblance. Most categories are not defined by necessary and sufficient conditions, but instead are like family resemblances. Individuals share graded membership in a category depending on the number of relevant attributes they have in common.

iv. Whewell on Types

v. Exactness

vi. Definitions and Rules

c. Critique of Logical Atomism

i. Limitations of Analysis

ii. Ideal Languages

2. Austin: Performatives

a. Background

b. Performatives. Many speech acts do not assert propositions; rather, they do things directly. Examples are "I promise . . . ," "I now pronounce you husband and wife."

c. Speech Act Theory

11.1.3 Images in Cognition

1. Rosch: Prototypes

2. Shepard: Images as Representations

Perhaps also Bruner and Kosslyn here.

3. Miller: Imagery in Scientific Thought

11.2 Field Theories

11.2.1 Development of Field Theory

a. Leibniz: Law of Continuity

b. The Contact Problem

c. Boscovich: Theory of Natural Philosophy

d. Faraday: Field Theory

11.2.2 Field Theory in Gestalt Psychology

a. Gestalts and Perception
b. Field Theory
c. Nalimov: Semantic Fields
d. Lewin: Field Theory in Sociology
e. The Failure of Gestalt Psychology

11.2.3 Pribram: Holonomic Brain Theory

1. Background

a. Freud's Project

b. Lashley. In 1929 Lashley ("In Search of the Engram") discovered that memory traces are not localized in the brain, but he could not imagine how they are represented. The answer is provided by holographic memory models.

2. Holographic Memory Models.

In 1963 P. J. van Heerden proposed the *holographic hypothesis*, which is the memory traces are represented in a distributed fashion analogous to optical holograms. A variety of holographic memory models, including optical and other implementations, were developed in the 1970s. These were a step toward later artificial neural network models (Sec. 11.5).

3. The Neural Wave Equation.

In 1966, neuroscientist Karl H. Pribram began to investigate the holographic hypothesis and developed it extensively over the following decades.

a. Dendritic Microprocesses
b. The Wave Equation (informally)

11.3 Phenomenology

11.3.1 Introduction

a. Romanticism
b. Hegel: Dialectical Process

11.3.2 Nietzsche: Trying Interpretations

In the following citations, S = Schacht (*N*), J = Jones (*HWP2*, Vol. IV).

a. Introduction Philology; inaccessibility of original text [J 235–6]

b. Perspectivism Inescapability of Interpretation, both in philosophy and science; plurality of drives [S 9, 61; J 236–8]

c. Trying Interpretations Tried, as by lawyers; cf. Sophists, & Popper's criticizability; experiment [S 5–7, 9, 18; J 241–3]

d. The Creative Philosopher's Task Criticizing the Background; An amplifier in the cooperative/competitive dynamic of ideas; the philosophical laborer is also necessary [S 14–5, 24, 41]

e. "Objectivity" Multiple Perspectives; Metaperspectives [S 9–10]

f. Higher Thinking Prudence + Science + Aesthetics [S 10]

g. Philosophical Prejudices Concepts, absolute knowledge, logic, reason[S 23, 49]

h. Philosophical Omissions Historical sense, psychological sense, knowledge of physiology [S 30–4]

i. Naturalist Epistemology Knowledge shaped by biology (survival), culture [S 52–55+]

j. Pitfalls of Language Primacy of verbs over nouns [S 29–30, 48–50; J 238–41]

k. Critique of Concepts [J 244]

l. Antisystematism Consistency as a sign of weakness [J 260–1]

m. The World Continuous, not structurally articulated; becoming, not being; flux, not structure; chaos, not order [S 61–2]

11.3.3 Dilthey: Naturwissenschaften vs. Geisteswissenschaften

11.3.4 Husserl

a. Definition of Phenomenology

b. "To the Things Themselves!"

c. Expectation

11.3.5 Heidegger: Being in the World

a. Introduction

b. Dasein

c. "Worlds"

d. Existence

e. Being-in

f. Ready-to-hand vs. Present-at-hand. A key point is that the present-at-hand presupposes the ready-to-hand.

g. Polanyi: Tacit Knowledge

h. "Care"

i. Being-ahead-of-itself

j. Understanding

k. Gadamer: Philosophical Hermeneutics

l. Zen and All That

11.3.6 Critique of Artificial Intelligence

a. Dreyfus: Critique of Cognitivism

b. Flores & Winograd: Computers and Cognition

11.4 The Embodied Mind

11.4.1 Kant: Theoretical and Practical Reason

11.4.2 Gestalt Philosophy of Mind

1. Phenomenal Realism & Psychophysical Isomorphism

2. Perception of Value and Meaning

3. Gibson: Ecological Approach

11.4.3 Jung: Archetypal Psychology

1. Archetypes of the Collective Unconscious

a. General

b. Archetypes as physical and mental

2. The Natural Numbers as Archetypes

a. General
b. One
c. Two
d. Three
e. Four

3. The "Unreasonable Effectiveness of Mathematics"

a. The Problem
b. Archetypes as an explanation
c. Archetypes of the continuum and the integers
d. Example: ancient atomism

11.4.4 Paying Attention to Nature

a. Naturalized Epistemology

b. Maturana: Autopoiesis and Cognition

c. Lakoff & Johnson: The Role of the Body in Cognition. A new view of categorization: *experiential realism.* Lakoff (1987). Lakoff & Johnson (1980) is an older collaboration dealing specifically with the way certain key metaphors structure our language and thought. Johnson (1987) deals more specifically with the role of the body in cognition. (EDIT add Phil in Flesh, L. & Nunez)

d. Sacks: Insights From Neuropathology. Sacks' description of "the man who mistook his wife for a hat" in the book of the same name illustrates how an AI would behave if it categorized objects in terms of abstract (e.g., geometric) predicates, without an embodied understanding. [Oliver Sacks, *The Man Who Mistook His Wife for a Hat*, Summit Books, 1985.]

e. Kuhn & Lakatos: The Sociology of Science

f. Popper: Evolutionary Epistemology

11.4.5 Artificial Situated Intelligence

a. Brooks: Subsumption Architectures

b. Artificial Life and Synthetic Ethology

Part IV. Knowledge in Flux

11.5 Neural Information Processing

11.5.1 Neural Information Processing in the Brain

1. Introduction

General introduction to the brain.

1. Properties of Neurons and Synapses

including dendritic computation

2. Localization of Brain Function

11.5.2 Artificial Neural Networks

1. History

a. McCulloch-Pitts Cells
b. Turing: B-type Networks
c. Rosenblatt: Perceptrons
d. Minsky & Papert Critique
e. Renaissance of Neural Nets

2. General Structure

a. Units and Connections
b. Layers
c. Learning & Training
i. Hardwired Connections
ii. Supervised Learning
iii. Reinforcement Learning
iv. Unsupervised Learning
v. Training vs. Programming
d. Distributed Knowledge Representation

3. Accomplishments

Short descriptions of a few successful neural net experiments with especial relevance to epistemology (mostly from Rumelhart & McClelland)

4. Differences from Biological Neural Nets

a. Axons are systematically connected
b. Synapses may have significant computational power
c. Spatio-temporal signal processing
d. Scale of parallelism

11.5.3 Continuous Computation

a. Analog and Digital Computation
b. The Decline of Analog Computation
c. An Analog Renaissance? The role of massively-parallel low-precision analog computation
d. Field Computers Optical Chemical Electronic Quantum

11.6 Continuous Cognition

11.6.1 Introduction

a. Is the discrete or the continuous more fundamental?
b. Complementarity Principle

11.6.2 Knowledge as a Structured Continuum

11.6.3 Continuous Formal Systems

11.6.4 Continuous Computability

a. For consistency, take a completely continuous view
b. All categories are graded
c. Computability in a continuous context

11.6.5 The Brain as a Dynamical Systems

a. State Space

b. Attractors

i. Simple

ii. Limit Cycles

iii. Quasiperiodic

iv. Chaotic

c. Implications for Reductionism

11.6.6 The Thermodynamics of Knowledge

11.7 Conclusions

Bibliography

Abbreviations

AHD	Morris (*AHD*)
DHI	Wiener (*DHI*)
DK	Diels & Kranz (*Frag.*)
EDM	Iyanaga & Kawada (*EDM*)
EP	Edwards (*EP*)
KRS	Kirk, Raven & Schofield (*Presoc.*)
LSJ	Liddell, Scott & Jones (*LSJ*)
OCD	Hammond & Scullard (*OCD*)

(Appiah, *A&C***).** A. Appiah, *Assertions and Conditionals*, Cambridge University Press, Cambridge UK, 1985.

(Austin, *PP***).** J. L. Austin, *Philosophical Papers*, 3rd ed., Edited by J. O. Urmson & G. J. Warnoch, Oxford: Oxford University Press, 1979.

(Ayer, *LP***).** Alfred Jules Ayer (Ed.), *Logical Positivism.* Free Press, New York, 1959.

(Ayer, *LTL***).** Alfred Jules Ayer, *Language, Truth and Logic*, second ed., Dover, New York, 1952.

(Barnes, *Aris.***).** Jonathan Barnes, *Aristotle*, Oxford University Press, Oxford, 1982. Only 101 pages, a good, short introduction to Aristotle's life and work.

(Barnes, *EGP***).** Jonathan Barnes, *Early Greek Philosophy*, Penguin Books, New York, 1987.

(Barwise & Etchemendy, *Liar***).** Jon Barwise and John Etchemendy, *The Liar: An Essay on Truth and Circularity*, Oxford University Press, New York and Oxford, 1987.

(Bellezza, *IYMS***).** Francis S. Bellezza, *Improve Your Memory Skills*, Prentice-Hall, Englewood Cliffs, 1982.

(Benacerraf & Putnam, *PM*). Paul Benacerraf and Hilary Putnam, *Philosophy of Mathematics: Selected Readings*, second edition, Cambridge University Press, Cambridge UK, 1983.

(Beth, *FM*). Evert W. Beth, *The Foundations of Mathematics: A Study in the Philosophy of Science*, Harper Torchbooks, New York, 1966.

(Beveridge, *ASI*). W. I. B. Beveridge, *The Art of Scientific Investigation*, Vintage Books, New York, 1950.

(Bishop, *FCA*). Errett Bishop, *Foundations of Constructive Analysis*, McGraw-Hill, New York, 1967.

(Blau, Cab). Cabala. In: Edwards (*EP*, Vol. 2, pp. 1–3)

(Boas, *MMPS*). Mary L. Boas, *Mathematical Methods in the Physical Sciences*, second edition, John Wiley & Sons, New York, 1983.

(Bocheński, *HFL*). I. M. Bocheński, *A History of Formal Logic*, Chelsea Publishing Co., New York, 1961, 1970. This book has a *very* complete bibliography.

(Bochner, *RMRS*). Salomon Bochner, *The Role of Mathematics in the Rise of Science*, Princeton University Press, Princeton, 1966.

(Bochner, Inf.). Salomon Bochner, Infinity. In: Wiener (*DHI*), Vol. II, pp. 604b–617a.

(Bolzoni, *GM*). Lina Bolzoni, *The Gallery of Memory: Literary and Iconographic Models in the Age of the Printing Press*, transl. by Jeremy Parzen, University of Toronto Press, Toronto, Buffalo & London, 2001.

(Boole, *ILT*). George Boole, *An Investigation of the Laws of Thought*, Macmillan, New York, 1854. Reprinted by Dover, New York, 1958.

(Boyer, IAG). Carl B. Boyer, The Invention of Analytic Geometry, in Morris Kline (ed.), *Mathematics: An Introduction to its Spirit and Use*, San Francisco CA, W. H. Freeman, 1979, pp. 39–44; originally published in *Scientific American*, January 1949.

(Brickhouse & Smith, *SoT*). Thomas C. Brickhouse & Nicholas D. Smith, *Socrates on Trial*, Princeton: Princeton University Press, 1989.

(Bronowski, *CSS*). Jacob Bronowski, *The Common Sense of Science*, Harvard University Press, Cambridge MA, 1978.

(Brown, *L1*). James Cooke Brown, *Loglan 1: A Logical Language*, third edition, The Loglan Institute, Gainseville, FL & Palm Springs, CA, 1975.

(Brown, *L4&5*). James Cooke Brown, *Loglan 4 & 5: A Loglan-English / English-Loglan Dictionary*, second ed., Loglan Institute, Gainseville FL & Palm Springs CA, 1975.

(Brown, Loglan). James Cooke Brown, Loglan, *Scientific American*, June 1960.

(Buchler, *PWP*). Justus Buchler (Ed.), *Philosophical Writings of Peirce*, Dover, New York, 1955. Reprint of *The Philosophy of Peirce: Selected Writings*, Routledge & Kegan Paul, London, 1940.

(Buridan, *Soph8*). Jean Buridan, *Sophismata. 8. capitulum*, transl. by G. E. Hughes, Cambridge University Press, Cambridge UK, 1982.

(Burkert, *LSAP*). Walter Burkert, *Lore and Science in Ancient Pythagoreanism*, Harvard University Press, Cambridge MA, 1972. Transl. by Edwin L. Minar, Jr. A comprehensive study; see esp. pp. 15–96, 369–482.

(Burnet, *GPI*). John Burnet, *Greek Philosophy Part I: Thales to Plato*, Macmillan, London, 1928.

(Burnet, *EGP*). John Burnet, *Early Greek Philosophy*, fourth edition, Macmillan, London, 1930.

(Butterworth, *AMC*). Charles E. Butterworth (trans., intro.), *Averroes' Middle Commentaries on Aristotle's* Categories *and* De Interpretatione, Princeton University Press, Princeton NJ, 1983.

(Bynum & al., *DHS*). W. F. Bynum, E. J. Browne & Roy Porter (eds.), *Dictionary of the History of Science*, Princeton University Press, Princeton NJ, 1981.

(Bywater, *Her.***).** I. Bywater, *Heracliti Ephesii reliquiae*, Oxford, 1877; reprinted in Richards (*Her.*).

(Canfora, *VL***).** Luciano Canfora, *The Vanished Library: A Wonder of the Ancient World*, transl. Martin Ryle, University of California Press, Berkeley CA, 1989.

(Carnap, LFUS). Rudolf Carnap, Logical Foundations of the Unity of Science. In Neurath, Carnap & Morris (*IEUS1*, pp. 42–62).

(Carnap, *LSL***).** Rudolf Carnap, Rudolf Carnap, *The Logical Syntax of Language*, tr. Amethe Smeaton (Countess von Zeppelin). Routledge & Kegan Paul, London, 1937.

(Cassuto, *CE***).** Umberto Cassuto, *A Commentary on the Book of Exodus*, 1st English ed. of *Perush 'al Sefer Shemot*. Magnes Press, Hebrew University, Jerusalem, 1967.

(Cherry, *OHC***).** Colin Cherry, *On Human Communication: A Review, A Survey and A Criticism*, second edition. Cambridge, MA: MIT Press, 1966.

(Childs, *Ex***).** Brevard S. Childs, *The Book of Exodus: A Critical, Theological Commentary.* Westminster Press, Philadelphia PA, 1974.

(Chomsky, *ATS***).** Noam Chomsky, *Aspects of the Theory of Syntax*, MIT Press, Cambridge, 1965.

(Chukovsky, *AT***).** Kornei Chukovsky, *The Art of Translation: Kornei Chukovsky's* A High Art, translated and edited by Lauren G. Leighton, University of Tennessee Press, Knoxville, 1984.

(Chuvin, *CLP***).** Pierre Chuvin, *A Chronicle of the Last Pagans*, transl. B. A. Archer, Harvard University Press, Cambridge MA, 1990.

(Clagett, *SMMA***).** Marshall Clagett, *The Science of Mechanics in the Middle Ages*, Oxford University Press, London, 1959.

(Clagett, RS). Marshall Clagett, Richard Swineshead and Late medieval Physics, *Osiris*, Vol. 9 (1950), pp. 131-161; reprinted in Marshall Clagett, *Studies in Medieval Physics and Mathematics*, Variorum Reprints, London, 1979, Ch. III.

(Clulee, *JDNP*). Nicholas H. Clulee, *John Dee's Natural Philosophy: Between Science and Religion*, Routledge, London, 1988.

(Cohen, *JBW*). David Cohen, *J. B. Watson: The Founder of Behaviourism*, Routledge & Kegan Paul, London, 1979.

(Cohen & Drabkin, *SBGS*). Morris R. Cohen & I. E. Drabkin, *A Source Book in Greek Science*, Harvard University Press, Cambridge MA, 1958.

(Cohen, *STCH*). Paul J. Cohen, *Set Theory and the Continuum Hypothesis*, W. A. Benjamin, New York, 1966.

(Cohen & Feigenbaum, *HoAI 3*). Paul R. Cohen & Edward A. Feigenbaum (eds.), *The Handbook of Artificial Intelligence, Volume III*, William Kaufmann, Los Altos, CA, 1982.

(Cornford, *FRP*). F. M. Cornford, *From Religion to Philosophy: A Study in the Origins of Western Speculation*, Atlantic Highlands, NJ: Humanities Press, 1912.

(Coudert, *L&K*). Allison P. Coudert, *Leibniz and the Kabbalah*, Kluwer, Dordrecht, 1995.

(Couturat, *Opl. Leib.*). Louis Couturat (ed.), *Opuscules et Fragments Inédits de Leibniz*, Georg Olms, Hildesheim, 1961.

(Crombie, *MEMS*). Alistair Cameron Crombie, *Medieval and Early Modern Science*, Doubleday, Garden City NY, 1959.

(Crystal, *CEL*). David Crystal, *The Cambridge Encyclopedia of Language*, Cambridge University Press, Cambridge, 1987.

(Curry, *OFPM*). H. B. Curry, *Outlines of a Formalist Philosophy of Mathematics*, North-Holland, Amsterdam, 1951.

(Curry, Feys & Craig, *CLI*). H. B. Curry, R. Feys & W. Craig, *Combinatory Logic, Vol. I*, Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1958.

(Davis, *ANSA*). Martin Davis, *Applied Nonstandard Analysis*, John Wiley & Sons, New York NY, 1977.

(Davis, *Undec.*). Martin Davis, *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Raven Press, Hewlett NY, 1965.

(Davis & Hersh, *ME*). Philip J. Davis & Reuben Hersh, *The Mathematical Experience*, Houghton-Mifflin, Boston, 1981.

(Dedekind, *ETN*). Richard Dedekind, *Essays on the Theory of Numbers*, transl. Wooster W. Beman, Open Court, La Salle IL, 1963.

(Dee, *MP*). John Dee, *The Mathematicall Praeface to the Elements of Geometrie of Euclid of Megara (1570)*, with intro. by Allen G. Debus, Science History Publications, New York NY, 1975.

(Dennett, *IS*). Daniel C. Dennett, *The Intentional Stance*, Cambridge, MA: The MIT Press, 1987.

(De Santillana, *CoG*). Giorgio De Santillana, *The Crime of Galileo*, University of Chicago Press, Chicago IL, 1955.

(Descartes, *Geom*). René Descartes, *The Geometry of René Descartes*, transl. by David Eugene Smith & Marcia L. Latham, Open Court, Chicago IL, 1925.

(Desjardins, *RE*). Rosemary Desjardins, *The Rational Enterprise: Logos in Plato's Theaetetus*, State University of New York Press, Albany, 1990.

(Diamond & Teichman, *I&I*). Cora Diamond & Jenny Teichman (Eds.), *Intention & Intentionality: Essays in Honour of G. E. M. Anscombe*, Cornell University Press, Ithaca NY, 1979.

(Diels & Kranz, *Frag.*). Hermann Diels, *Die Fragmente der Vorsokrater*, tenth edition, revised with additions and index by Walther Kranz, Weidmannsche Verlagsbuchhandlung, Berlin, 1952.

(Dijksterhuis, *MWP*). E. J. Dijksterhuis, *The Mechanization of the World Picture: Pythagoras to Newton*, transl. C. Dikshoorn, Princeton University Press, Princeton NJ, 1961.

(Dobbs, *FNA*). B. J. T. Dobbs, *The Foundation of Newton's Alchemy; or, "The Hunting of the Greene Lyon."* Cambridge University Press, Cambridge UK, 1975.

(Dobbs, NCET). B. J. T. Dobbs, Newton's *Commentary* on the *Emerald Tablet* of Hermes Trismegistus: Its Scientific and Theological Significance. In: Merkel & Debus (*H&R*, pp. 182–191).

(Dodds, *GI*). E. R. Dodds, *The Greeks and the Irrational*, University of California Press, Berkeley, 1951.

(Donnegan, *Lex.*). James Donnegan, *A New Greek and English Lexicon*, J. F. Dove, London, 1831.

(Douglas, *NBD*). James Dixon Douglas, *The New Bible Dictionary*, 1st ed. Eerdmans, Grand Rapids MI, 1962.

(Dreyfus, 1982). Hubert L. Dreyfus, Introduction, In Dreyfus & Harrison (*HICS*, pp. 1-27).

(Dreyfus & Harrison, *HICS*). Hubert L. Dreyfus, with Harrison Hall (eds.), *Husserl, Intentionality, and Cognitive Science*, The MIT Press, Cambridge MA, 1982.

(Durant, *Civ 7*). Will & Ariel Durant, *The Story of Civilization: Part VII, The Age of Reason Begins*, Simon & Schuster, New York NY, 1961.

(Dusek, *HIP*). Val Dusek, *The Holistic Inspirations of Physics: The Underground History of Electromagnetic Theory*, Rutgers University Press, New Brunswick, 1999.

(Eco, *SPL*). Umberto Eco, *The Search for the Perfect Language*, transl. James Fentress, Cambridge, MA: Blackwell, 1997.

(Edwards, *EP*). Paul Edwards, *The Encyclopedia of Philosophy*, Macmillan Co. & The Free Press, New York, 1967.

(Ellegård, 1973). Alvar Ellegård, Language, Study of. In: Wiener (*DHI*), pp. 659b–673b.

(Englefield, *Lang.*). Ronald Englefield, *Language: Its Origin and Relation to Thought*, edited by G. A. Wells & D. R. Oppenheimer, New York: Charles Scribner's Sons, 1977.

(Estes, GW). Leland L. Estes, Good Witches, Wise Men, Astrologers, and Scientists: William Perkins and the Limits of the European Witch-Hunts. In: Merkel & Debus (*H&R*, pp. 154–165).

(Flew, *DP*). Antony Flew, *A Dictionary of Philosophy*, revised second edition, St. Martins Press, New York, 1984.

(Fodor, *LT*). J. A. Fodor, *The Language of Thought*, Crowell, New York, 1975.

(Fodor, *Rep.*). J. A. Fodor, *Representations*, MIT Press, Cambridge MA, 1981.

(Fodor, *PS*). J. A. Fodor, *Psychosemantics*, MIT Press, Cambridge MA, 1987.

(Frazer, *GB*). James George Frazer, *The Golden Bough: A Study in Magic and religion*, 1 vol. abridged edition, New York: Macmillan, 1922.

(Frazer, *NGB*). James George Frazer, *The New Golden Bough: A New Abridgement of the Classic Work by Sir James George Frazer*, edited by Theodor H. Gaster, New York: Criterion Books, 1959.

(Freddoso & Schuurman, *OTP*). Alfred J. Freddoso & Henry Schuurman (transl.), *Ockham's Theory of Propositions: Part II of the Summa Logicae*, Notre Dame: University of Notre Dame Press, 1980.

(Freeman, *APSP*). Kathleen Freeman, *Ancilla to the Pre-Socratic Philosophers: A Complete Translation of the Fragments in Diels, Fragmente der Vorsokrater*, Harvard University Press, Cambridge MA, 1948.

(French, *JD*). Peter J. French, *John Dee: The World of an Elizabethan Magus*, Routledge & Kegan Paul, London, 1972.

(Gabor, TC). Dennis Gabor, Theory of Communication, *Journal of the Institute of Electrical Engineers*, Vol. 93 Part III (1946), pp. 429–457.

(Gardner, *LMD*). Martin Gardner, *Logic Machines and Diagrams*, second edition, University of Chicago Press, Chicago, 1958, 1982.

(Gardner, *MNS*). Howard Gardner, *The Mind's New Science: A History of the Cognitive Revolution*, Basic Books, New York, 1985.

(Glare, *OLD*). P. G. W. Glare, *Oxford Latin Dictionary*, Clarendon Press, Oxford, 1982.

(Gödel, WCCP). Kurt Gödel, What is Cantor's Continuum Problem? In: Benacerraf & Putnam (*PM*, pp. 470–485).

(Grant, *PSMA*). E.? Grant, *Physical Science in the Middle Ages*, Cambridge University Press, Cambridge UK, 1971/1977.

(Goodman, *LA*). Nelson Goodman, *Languages of Art: An Approach to a Theory of Symbols*, Bobbs-Merrill, Indianapolis IN & New York NY, 1968.

(Goudge, *ToP*). Thomas A. Goudge, *The Thought of C. S. Peirce*, Dover, New York, 1969. Originally published by University of Toronto Press, Toronto, 1950.

(Gould, *MM*). Stephen Jay Gould, *The Mismeasure of Man*, W. W. Norton, New York, 1981.

(Gregory, *OCM*). Richard L. Gregory with O. L. Zangwill (eds.), *The Oxford Campanion to the Mind*, Oxford University Press, Oxford UK, 1987.

(Griswold, SGP). Charles L. Griswold, Jr. Stoning Greek Philosophers: Platonic Political Philosophy and the Trial of Socrates, *Classical Bulletin*, **67** (1991), No. 1, pp. 3–15.

(Guirand, *NLEM*). Félix Guirand, *New Larousse Enclyclopedia of Mythology*, translated and revised by Richard Aldington, Delano Ames, et al., Crescent Books, New York, 1968.

(Guthrie, *HGP*). W. K. C. Guthrie, *A History of Greek Philosophy*, Vols. I, II, III, Cambridge University Press, Cambridge UK, 1962, 1965, 1969.

(Halmos, *LBA*). Paul R. Halmos, *Lecures on Boolean Algebras*, D. Van Nostrand, Princeton NJ, 1963.

(Hamilton & Cairns, *CDP*). Edith Hamilton and Huntington Cairns (eds), *The Collected Dialogues of Plato*, Princeton University Press, Princeton, 1961.

(Hamlyn, HoE). D. W. Hamlin, "Epistemology, History of," in *EP*, Vol. 3, pp. 3–38.

(Hamming, *C&IT*). Richard W. Hamming, *Coding and Information Theory*, Prentice-Hall, Englewood Cliffs, 1980.

(Hamming, UEM). R. W. Hamming, "The Unreasonable Effectiveness of Mathematics," *American Mathematical Monthly*, Vol. 87, No. 2 (February 1980), pp. 81–90.

(Hammond & Scullard, *OCD*). N. G. L. Hammond and H. H. Scullard, *The Oxford Classical Dictionary*, Oxford University Press, Oxford, 1970.

(Hardy, MA). G. H. Hardy, "A Mathematician's Apology," Newman (*WM*, Vol. 4, pp. 2027–2038).

(Harper & al., *Ifs*). W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*, Reidel, Dordrecht Holland, 1981.

(Hartnack, *W&MP*). Justus Hartnack, *Wittgenstein and Modern Philosophy,* second ed., tr. Maurice Cranston, University of Notre Dame Press, Notre Dame, 1986.

(Hausdorff, *ST*). Felix Hausdorff, *Set Theory*, translated by John R. Aumann et al., Chelsea, New York, 1957.

(Heath, *Euclid*). Euclid, *The Thirteen Books of Euclid's Elements*, Dover, New York, 1956. Translated by T. L. Heath. A definitive translation of the *Elements*, with extensive commentary (three volumes).

(Hastings, *DB*). James Hastings, *Dictionary of the Bible*, revised ed. by Frederick C. Grant and H. H. Rowley. Scribner, New York, 1963.

(Heidegger, *EGT***).** Martin Heidegger, *Early Greek Thinking: The Dawn of Western Philosophy*, second edition, tr. David Farrell Krell and Frank A. Capuzzi, Harper & Row, San Francisco, 1984.

(Heisenberg, *P&P***).** Werner Heisenberg, *Physics and Philosophy: The Revolution in Modern Science*, Harper & Brothers, New York, 1958.

(Heytesbury, *OIS***).** William Heytesbury, *On "Insoluble" Sentences: Chapter One of his Rules for Solving Sophisms*, transl. by Paul Vincent Spade, Pontifical Institute of Mediaeval Studies, Toronto, 1979.

(Hindemith, *CMC***).** Paul Hindemith, *The Craft of Musical Composition, Book I: Theoretical Part*, fourth edition, tr. Arthur Mendel, Schott, New York, 1942.

(Hobbes, *Lev.***).** Thomas Hobbes, *Leviathan, or The Matter, Forme, & Power of a Common-Wealth Ecclesiasticall and Civill*, Crooke, London, 1651. Penguin edition, New York, 1968.

(Hughes, *JBSR***).** G. E. Hughes, *John Buridan on Self-Reference*, Cambridge University Press, Cambridge UK, 1982.

(Humez & Humez, *ABC***).** Alexander Humez and Nicholas Humez, *ABC et cetera: The Life & Times of the Roman Alphabet*, David R. Godine, Boston, 1985.

(Hurd & Loeb, *INSRA***).** Albert E. Hurd & Peter A. Loeb, *An Introduction to Nonstandard Real Analysis*, Orlando FL, Academic Press, 1985.

(Hussey, *Presoc.***).** Edward Hussey, *The Presocratics*, Duckworth, London, 1972.

(Idel, *AP***).** Moshe Idel, *Absorbing Perfections : Kabbalah and Interpretation*, Yale University Press, New Haven, 2002.

(Idel, *K***).** Moshe Idel, *Kabbalah: New Perspectives*, Yale University Press, New Haven & London, 1988.

(Iyanaga & Kawada, *EDM***).** S. Iyanaga and Y. Kawada, *Encyclopedic Dictionary of Mathematics*, MIT Press, Cambridge MA, 1977.

(Jackson, *Cond*). F. Jackson, *Conditionals*, Basil Blackwell, Oxford UK, 1987.

(Jantsch, *SOU*). Erich Jantsch, *The Self-Organizing Universe: Scientific and Human Implications of the Emerging Paradigm of Evolution*, Pergamon, Oxford, 1980.

(Jensen & Wirth, *PUMR*). K. Jensen, N. Wirth, A. B. Mickel and J. F. Miner, *Pascal User Manual and Report*, third edition, Springer-Verlag, Berlin, 1985.

(Jevons, *ELL*). W. Stanley Jevons, *Elementary Lessons in Logic: Deductive and Inductive*, New York, NY: Macmillan, 1894.

(Jevons, OMPLI). W. Stanley Jevons, On the Mechanical Performance of Logical Inference. *Philosophical Transactions*, Vol. 160 (1870), p. 497; reprinted in Jevons, *Pure Logic and Minor Works* (1890).

(Jevons, *PS*). W. Stanley Jevons, *The Principles of Science: A Treatise on Logic and Scientific Method*, second ed., New York, NY: Dover, 1958.

(Johnson, *BiM*). Mark Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press, 1987.

(Johnston, *SLRL*). Mark D. Johnston, *The Spiritual Logic of Ramon Llull*. Oxford: Clarendon Press, 1987.

(Jonas, Gn). Hans Jonas, Gnosticism. In: Edwards (*EP*, Vol. 3, pp. 336–342)

(Jones, *Hipp.*). W. H. S. Jones, *Hippocrates, Vol. IV*, Loeb Classical Library Vol. 150, Harvard University press, Cambridge Mass., 1931. Includes Heraclitus' *On the Universe*.

(Jones, *HWP*). W. T. Jones, *A History of Western Philosophy*, 1st ed., Harcourt Brace Jovanovich, San Diego CA, 1969.

(Jones, *HWP2*). W. T. Jones, *A History of Western Philosophy*, 2nd ed., 5 vols., Harcourt Brace Jovanovich, New York NY, 1975.

(Joseph, ***IL*****).** H. W. B. Joseph, *An Introduction to Logic*, Oxford University Press, Oxford, 1916. A comprehensive presentation of traditional logic; by no means an introduction.

(Kahn, ***Be*****).** Charles H. Kahn, *The Verb 'Be' in Ancient Greek*, Dordrecht, 1973.

(Kahn, ***ATH*****).** Charles H. Kahn, *The Art and Thought of Heraclitus: An Edition of the Fragments with Translation and Commentary*, Cambridge University Press, Cambridge UK, 1979.

(Kaufmann, ***PC*****).** Walter Kaufmann, *Philosophical Classics*, Prentice-Hall, Englewood Cliffs, 1961. Two volumes. Basic texts with prefaces.

(Keisler, ***EC*****).** H. Jerome Keisler, *Elementary Calculus*, Prindle, Weber and Schmidt, Boston MA, 1976.

(Keisler, ***FIC*****).** H. Jerome Keisler, *Foundation of Infinitesmal Calculus*, Prindle, Weber and Schmidt, Boston MA, 1976.

(Kepler, ***GB*****).** Johannes Kepler, *Epitome of Copernican Astronomy*: sc iv and v. *The Harmonies of the World*: v, tr. by Charles Glenn Wallis, Great Books of the Western World, Vol. 16, Encyclopaedia Britannica, Chicago IL, 1952.

(Kerényi, ***El*****).** Carl Kerényi, *Eleusis: Archetypical Image of Mother and Daughter*, transl. by Ralph Manheim. Princeton University Press, Princeton NJ, 1967.

(Kerferd, ***SM*****).** George Briscoe Kerferd, *The Sophistic Movement*, Cambridge University Press, Cambridge, 1981.

(Kerferd, ***SATL*****).** George Briscoe Kerferd (ed.), *The Sophists and Their Legacy*, Franz Steiner Verlag, Wiesbaden, 1981.

(Kirk, Raven & Schofield, ***Presoc.*****).** G. S. Kirk, J. E. Raven & M. Schofield, *The Presocratic Philosophers*, second edition, Cambridge University Press, Cambridge UK, 1984. The first edition (1957) is more outspoken in its conclusions.

(Kline, ***MT*****).** Morris Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York NY, 1972.

(Kneale & Kneale, *DL*). William Kneale and Marthe Kneale, *The Development of Logic*, Oxford University Press, Oxford, 1962.

(Kneebone, *MLFM*). G. T. Kneebone, *Mathematical Logic and the Foundations of Mathematics*, Van Nostrand, London, 1963.

(Lakatos, *PR*). Imre Lakatos, *Proofs and Refutations: The Logic of Mathematical Discovery*, edited by John Worrall and Elie Zahar, Cambridge University Press, Cambridge UK, 1976.

(Lakoff, *WFDT*). George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, Chicago, 1987.

(Lakoff & Johnson, *MWLB*). George Lakoff and Mark Johnson, *Metaphors We Live By*, University of Chicago Press, Chicago, 1980.

(Laotse, *WoL*). Laotse, *The Wisdom of Laotse*, transl. & ed. by Lin Yutang, Modern Library, New York, 1948.

(Large, *ALM*). Andrew Large, *The Artificial Language Movement*, Basil Blackwell, New York, 1985.

(Lecky, *RIRE*). William Edward Hartpole Lecky, *History of the Rise and Influence of the Spirit of Rationalism in Europe*, Watts & Co., London, 1910.

(Leibniz, Academy). G. W. Leibniz, *Gottfried Wilhelm Leibniz: Sämtliche Schriften und Briefe*, Academy Edition, Darmstadt and Berlin, 1923–.

(Leibniz, Couturat). G. W. Leibniz, *Opuscules et Fragments inédits de Leibniz*, edited by L. Couturat, Paris, 1903.

(Lemprière, *Cl. Dict.*). J. Lemprière, *Lemprière's Classical Dictionary of Proper Names Mentioned in Ancient Authors*, revised by F. A. Wright, Routledge & Kegan Paul, London, 1949.

(Liddell, Scott & Jones, *LSJ*). Henry George Liddell & Robert Scott, *A Greek-English Lexicon*, ninth edition, revised and augmented by Henry Stuart Jones, with a supplement (ed. by E. A. Barber), Oxford University Press, Cambridge UK, 1968.

(Llull, *DI*). Ramon Llull, *Doctor Illuminatus: A Ramon Llull Reader*, ed. & transl., Anthony Bonner, Princeton, NJ: Princeton Univ. Press, 1993.

(Llull, *SWRL*). Ramon Llull, *Selected Works of Ramon Llull (1232–1316)*, Vol. I, edited & transl. by Anthony Bonner, Princeton, NJ: Princeton University Press, 1985.

(Long & Sedley, 1987). A. A. Long & D. N. Sedley, *The Hellenistic Philosophers: Vol. 1, Translations of the Principal Sources with Philosophical Commentaries*, Cambridge, UK: Cambridge University Press, 1987.

(Loux, *OTT*). Michael J. Loux (transl.), *Ockham's Theory of Terms: Part I of the Summa Logicae*, Notre Dame: University of Notre Dame Press, 1974.

(Lovejoy, *GCB*). Arthur O. Lovejoy, *The Great Chain of Being: A Study of the History of an Idea*, Harvard University Press, Cambridge & London, 1964.

(Lutz, *NSA*). Robert Lutz, *Nonstandard Analysis: A Practical Guide with Applications*, Berlin, New York NY, Springer-Verlag, 1981.

(MacLane & Birkhoff, *Alg*). S. MacLane & G. Birkhoff, *Algebra*, Macmillan, New York, 1967.

(MacLennan, CCKR). Bruce J. MacLennan, Characteristics of Connectionist Knowledge Representation, *Information Sciences*, Vol. 70 (1993), pp. 119–143.

(MacLennan, CSS). Bruce J. MacLennan, Continuous Symbol Systems: The Logic of Connectionism, *Neural Networks for Knowledge Representation and Inference*, edited by Daniel S. Levine and Manuel Aparicio IV, Hillsdale, NJ: Lawrence Erlbaum, 1993, pp. 83–120.

(MacLennan, DD). Bruce J. MacLennan, The Discomforts of Dualism *Behavioral and Brain Sciences* Vol. 13, No. 4 (December 1990), pp. 673–674. A review of Penrose (*ENM*).

(MacLennan, *FP*). Bruce J. MacLennan, *Functional Programming: Practice and Theory*, Addison-Wesley, Reading, 1990.

(MacLennan, GAC). Bruce J. MacLennan, Grounding Analog Computers, *Think*, in press. This is a commentary on S. Harnad, "Grounding Symbols in the Analog World with Neural Nets."

(MacLennan, IS). Bruce J. MacLennan, Image and Symbol: Continuous Computation and the Emergence of the Discrete, *Artificial Intelligence and Neural Networks: Steps Toward Principled Integration, Volume I: Basic Paradigms; Learning Representational Issues; and Integrated Architectures*, edited by Vasant Honavar and Leonard Uhr, New York, NY: Academic Press, in press.

(MacLennan, LNAI). Bruce J. MacLennan, Logic for the New AI. James H. Fetzer (ed.), *Aspects of Artificial Intelligence*, Kluwer, Dordrecht, 1988, pp. 163–192.

(MacLennan, ODFS). Bruce J. MacLennan, "Observations on the Differences Between Formulas and Sentences and their Application to Programming Language Design," *SIGPLAN Notices* **14**, 7 (July 1979), pp. 51–61.

(MacLennan, *POPL*). Bruce J. MacLennan, *Principles of Programming Languages: Design, Evaluation, and Implementation*, third edition, Oxford University Press, New York & Oxford, 1999.

(MacLennan, PV). Bruce J. MacLennan, Predicate Variables, *The Loglanist*, Vol. 1, No. 3 (February 1977), pp. 172–173.

(MacLennan, WLIOW). Bruce J. MacLennan, Words Lie in Our Way, *Minds and Machines*, in press.

(Maier, *OTES*). Anneliese Maier, *On the Threshold of Exact Science: Selected Writings of Anneliese Maier on Late Medieval Natural Philosophy*, edited and transl. by Steven D. Sargent, University of Pennsylvania Press, Philadelphia, 1982.

(Mallory, *SIE*). J. P. Mallory, *In Search of the Indo-Europeans*, Thames and Hudson, London, 1989.

(Markov, *TA*). A. A. Markov, *Theory of Algorithms* (US Dept. of Commerce Office of Technical Services OTS 60-51085), transl. by Jacques

J. Schorr-Kon and PST Staff, Israel Program for Scientific Translation, Jerusalem, 1961. Translation of *Teoriya Algorifmov*, Academy of Sciences of the USSR, Moscow, 1954.

(Martin, *PL***).** Robert L. Martin (ed.), *The Paradox of the Liar*, Yale University Press, New Haven and London, 1970.

(Mays & Henry, J&L). W. Mays & D. P. Henry, Jevons and Logic. *Mind*, Vol. 62 (October 1953), p. 484.

(Maziarz & Greenwood, *GMP***).** Edward A. Maziarz and Thomas Greenwood, *Greek Mathematical Philosophy*, Frederick Ungar, New York, 1968. See Chs. 2–5 on Pythagoreanism, Ch. 6 on Zeno. Part IV is devoted to Euclid.

(McDaniel & Pressley, *IRMP***).** Mark A. McDaniel & Michael Pressley (eds.), *Imagery and Related Mnemonic Processes: Theories, Individual Differences, and Applications*, Springer-Verlag, New York, 1987.

(McKeon, *SMP***).** Richard McKeon (ed. & transl.), *Selections from Medieval Philosophers* (2 vols.), Charles Scribner's, New York, 1929/1930.

(Mendelson, *IML***).** Elliot Mendelson, *Introduction to Mathematical Logic*, Van Nostrand Reinhold, New York, 1964.

(Merkel & Debus, *H&R***).** Ingrid Merkel & Allen G. Debus, *Hermeticism and the Renaissance: Intellectual History and the Occult in Early Modern Europe*, Folger Books, Washington DC, 1988.

(Mollett, *Dict.***).** J. W. Mollett, *An Illustrated Dictionary of Ancient Art and Archaeology*, Omega Books, London, 1987; originally published as *An Illustrated Dictionary of Words Used in Art and Archaeology*, Sampson, Low, Marston, Searle and Rivington, London, 1883.

(Moody, ML). Ernest A. Moody, Medieval Logic. In: "Logic, History of" In: Edwards (*EP*), Vol. 4, pp. 528–534.

(Moser, GEB). David Moser, The translation of *Gödel, Escher, Bach* into Chinese. *Babel*, in press.

(Morris, *AHD*). William Morris, *The American Heritage Dictionary of the English Language*, Boston: Houghton Mifflin, 1978.

(Nahm, *SEGP*). Milton C. Nahm (ed.), *Selections From Early Greek Philosophy*, third edition, F. S. Crofts & Co., New York, 1947. A classic source book (4 volumes) in readable mathematics, recently reprinted.

(Nagel & Newman, GP/WM). Ernest Nagel and James R. Newman, Goedel's Proof. In: Newman (*WM*), pp. 1668–1695.

(Nagel & Newman, *GP*). Ernest Nagel and James R. Newman, *Gödel's Proof*, New York University Press, New York, 1958.

(Nalimov, *ILL*). V. V. Nalimov, *In the Labyrinths of Language: A Mathematician's Journey*, ed. by Robert G. Colodny, ISI Press, Philadelphia PA, 1981.

(Nettleship & Sandys, *Dict.*). Henry Nettleship and J. E. Sandys, *A Dictionary of Classical Antiquities: Mythology, Religion, Literature and Art*, sixth edition, Swan Sonnenschein and Co., London, 1901.

(Neugebauer, *ESA*). Otto Neugebauer, *The Exact Sciences in Antiquity*, 2nd ed., Brown University Press, Providence RI, 1957.

(Neurath, PS). Otto Neurath, Protocol Sentences, tr. G. Schick. In Ayer (*LP*, pp. 199–208).

(Neurath, Carnap & Morris, *IEUS1*). Otto Neurath, Rudolf Carnap & Charles Morris, *International Encyclopedia of Unified Science I*, University of Chicago Press, Chicago, 1955.

(Newman, *WM*). James R. Newman, *The World of Mathematics*, Simon and Schuster, New York, 1956.

(Nietzsche, *WP*). Friedrich Nietzsche, *The Will to Power*, tr. Walter Kaufmann & R. J. Hollingdale, ed. Walter Kaufmann, Vintage Books, New York, 1967.

(Nietzsche, *BG&E*). Friedrich Nietzsche, *Beyond Good and Evil: Prelude to a Philosophy of the Future,* tr. Walter Kaufmann, Vintage, New York, 1966.

(Ogden & Richards, *MoM***).** C. K. Ogden & I. A. Richards, *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*, eighth edition, New York: Harcourt, Brace & World, 1946.

(Onians, *OET***).** Richard Broxton Onians, *The Origins of European Thought about the Body, the Mind, the Soul, the World, Time and Fate.*

(Otto, *HG***).** Walter F. Otto, *The Homeric Gods: The Spiritual Significance of Greek Religion*, tr. by Moses Hadas, Pantheon, New York NY, 1954. Cambridge, UK: Cambridge University Press, 1951.

(Pagel, *Par***).** W. Pagel, *Paracelsus: An Introduction to Philosophical Medicine in the Era of the Renaissance*, Bâle, New York NY, 1958.

(Parkinson, *LLP***).** G. H. R. Parkinson, *Leibniz Logical Papers*, Oxford University Press, Oxford, 1966. A very handy collection of Leibniz's writings on knowledge representation and logical calculi.

(Patrick, *Her.***).** G. T. W. Patrick, *The Fragments of the Work of Heraclitus of Ephesus: On Nature*, Baltimore, 1889; reprinted in Richards (*Her.*).

(Peirce, *CP***).** Charles Sanders Peirce, *Collected Papers of Charles Sanders Peirce* (8 vols.), C. Hartshorne, P. Weiss & A. W. Burks (eds.), Harvard University Press, Cambridge, 1931–1958.

(Penrose, *ENM***).** Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford University Press, Oxford, 1989.

(Peters, *GPT***).** F. E. Peters, *Greek Philosophical Terms*, New York University Press, New York, 1967. Contains analyses of the evolution of key philosophical terms through their use by various philosophers.

(Poncé, *K***).** Poncé, Charles, *Kabbalah: An Introduction and Illumination for the World Today*, Quest Books, San Francisco, CA, 1973.

(Popper, *OS&E***).** Karl R. Popper, *The Open Society and Its Enemies*, Princeton University Press, Princeton, 1950.

(Popper, *C&R*). Karl R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Basic Books, New York, 1962.

(Pour-El & Richards, 1979). Pour-El, M. B., & Richards, I., A Computable Ordinary Differential Equation Which Possesses No Computable Solution, *Annals of Mathematical Logic*, Vol. 17 (1979), pp. 61–90.

(Pour-El & Richards, 1981). Pour-El, M. B., & Richards, I., The Wave Equation with Computable Initial Data Such That its Unique Solution is Not Computable, *Advances in Mathematics*, Vol. 39 (1981), pp. 215–239.

(Pour-El & Richards, 1982). Pour-El, M. B., & Richards, I., Noncomputability in Models of Physical Phenomena, *International Journal of Theoretical Physics*, Vol. 21 (1982), pp. 553–555.

(Pribram, *B&P*). Karl H. Pribram, *Brain and Perception: Holonomy and Structure in Figural Processing*, Lawrence Erlbaum, Hillsdale NJ, 1991.

(Prier, *AL*). Raymond Adolph Prier, *Archaic Logic: Symbol and Structure in Heraclitus, Parmenides, and Empedocles*, Mouton, The Hague, 1976.

(Prigogine & Stengers, *OOC*). Ilya Prigogine and Isabelle Stengers, *Order Out of Chaos: Man's New Dialogue With Nature*, Bantam Books, New York, 1984.

(Prior, Bur). A. N. Prior, Some Problems of Self-reference in John Buridan, in A. N. Prior, *Papers in Logic and Ethics* (ed. P. T. Geach & A. J. P. Kenny), Duckworth, London UK, 1976; originally published in *Proceedings of the British Academy*, Vol. 48 (1962), pp. 281–296.

(Putnam, *MLR*). Hilary Putnam, *Mind, Language and Reality: Philosophical Papers*, Cambridge University Press, Cambridge, 1975.

(Randall, *Aris.*). John Herman Randall, Jr., *Aristotle*, Columbia University Press, New York, 1960. A good general introduction to Aristotle's philosophy.

(Rankin, OEA). H. D. Rankin, Ouk Estin Antilegein. In: Kerferd (*SATL*), pp. 25–37.

(Rescher, LILC). Nicholas Rescher, Leibniz's Interpretation of his Logical Calculi. *Journal of Symbolic Logic*, Vol. 19, No. 1 (March 1954), 1–13.

(Rescher, *PPS***).** Nicholas Rescher, *Peirce's Philosophy of Science: Critical Studies in the Theory of Induction and Scientific Method*, University of Notre Dame Press, Notre Dame, 1978.

(Reynolds & Wilson, *S&S***).** L. D. Reynolds and N. G. Wilson, *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature*, second edition, Oxford University Press, Oxford, 1974.

(Rheingold, *THWFI***).** Howard Rheingold, *They Have a Word for It: A Lighthearted Lexicon of Untranslatable Words and Phrases*, Jeremy P Tarcher, Los Angeles CA, 1988.

(Richards, *Her.***).** Lewis A. Richards (ed.), *Heraclitus of Ephesus*, Argonaut, Chicago, 1969. This binds in one volume Patrick (*Her.*) and Bywater (*Her.*) together with an introduction and bibliography by Richards.

(Rice, CRES). H. G. Rice, Classes of Recursively Enumerable Sets and Their Decision Problems, *Transactions of the American Mathematical Society*, Vol. 74, pp. 358–366.

(Robinson, *IEGP***).** John Mansley Robinson, *An Introduction to Early Greek Philosophy*, Houghton Mifflin Co., Boston, 1968. The chief fragments and ancient testimony, with connecting commentary.

(Robinson, *CA***).** T. M. Robinson, *Contrasting Arguments: An Edition of the* Dissoi Logoi, Arno Press, New York, 1979.

(Robinson, *NSA***).** Abraham Robinson, *Nonstandard Analysis*, North Holland, Amsterdam, 1966.

(Ronan, *Science***).** Colin A. Ronan, *Science: Its History and Development Among the World's Cultures*, Facts on File, New York, 1982.

(Rosser, ESTGC). J. B. Rosser, Extension of Some Theorems of Gödel and Church, *Journal of Symbolic Logic*, Vol. 1 (1936), pp. 87–91.

(Rossi, *LAM*). Paolo Rossi, *Logic and the Art of Memory: The Quest for a Universal Language*, transl. & intro., Stephen Clucas. Univ. of Chicago press, Chicago, 2000.

(Runes, *Dict.*). Dagobert D. Runes (ed.), *The Dictionary of Philosophy*, Philosophical Library, New York, 1942.

(Runes, *ToP*). Dagobart D. Runes (ed.), *Treasury of Philosophy*, Philosophical Library, New York, 1955.

(Russell, *PM*). Bertrand Russell, *The Principles of Mathematics*, W. W. Norton & Company, New York, 1903.

(Russell, *IMP*). Bertrand Russell, *Introduction to Mathematical Philosophy*, Macmillan, New York, 1919.

(Russell, *HWP*). Bertrand Russell, *A History of Western Philosophy*, Simon and Schuster, New York NY, 1943.

(Russell, LA). Bertrand Russell, Logical Atomism, in Ayer (*LP*, pp. 31–51).

(Russell, OD). Russell, Bertrand. On Denoting, *Logic and Knowledge: Essays 1901–1950*, ed. C. Marsh, Allen & Unwin, London, 1956.

(Russell, *PLA*). Bertrand Russell, *The Philosophy of Logical Atomism*, ed. David Pears, Open Court, La Salle IL, 1985; originally published 1918; previously published as *Logical Atomism* (1924) and *Russell's Logical Atomism* (Fontana, London, 1972).

(Russell & Whitehead, *PM*). Bertrand Russell and Alfred North Whitehead, *Principia Mathematica*, Cambridge University Press, Cambridge UK, 1910.

(Schacht, *N*). Richard Schacht, *Nietzsche*, Routledge & Kegan Paul, London, 1983.

(Schlick, FK). Moritz Schlick, The Foundations of Knowledge, tr. D. Rynin. In Ayer (*LP*, 209–227).

(Schlick, P&R). Moritz Schlick, Positivism and Realism, tr. D. Rynin. In Ayer (*LP*, 82–107).

(Schmandt-Besserat, ARS). D. Schmandt-Besserat, An Archaic Recording System and the Origin of Writing. *Syro-Mesopotamian Studies*, Vol. 1, No. 2 (July 1977).

(Schmandt-Besserat, EPW). D. Schmandt-Besserat, The Earliest Precursor of Writing. In Wang (*LWC*), pp. 31–40; reprinted from *Scientific American*, June 1978.

(Scholem, *Kab*). Gershom Scholem, *Kabbalah*, Quadrangle/The New York Times Book Co., New York NY, 1974.

(Scholem, *K&S*). Scholem, Gershom, *On the Kabbalah and Its Symbolism*, Schocken, New York, NY, 1960.

(Scholem, *Z*). Scholem, Gershom, *Zohar — The Book of Splendor: Basic Readings from the Kabbalah*, Schocken, New York, NY, 1963.

(Schwartz, *WTAC*). Schwartz, Benjamin I., *The World of Thought in Ancient China*, Harvard University Press, Cambridge, MA, 1985.

(Searle, *Int*). John R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge UK, 1983.

(Shannon, SARSC). Claude Shannon, A Symbolic Analysis of Relay and Switching Circuits, Master's Thesis, Massachusetts Institute of Technology; published in *Transactions of the American Institute of Electrical Engineers* **57** ($\boxed{\text{date?}}$), pp. 1–11.

(Sinnige, *M&I*). Theo Gerard Sinnige, *Matter and Infinity in the Presocratic Schools and Plato*, VanGorcum, Assen, 1968. See especially Chapter III. Good discussion of presocratic ideas of chaos and the infinite as indeterminate.

(Skinner, BSO). B. F. Skinner, Behaviourism, Skinner On. In *OCM*, pp. 74–75.

(Small, *WTM*). Jocelyn Penny Small, *Wax Tablets of the Mind: Cognitive Studies of Memory and Literacy in Classical Antiquity*, Routledge, London & New York, 1997.

(Smith, *HM*). D. E. Smith, *History of Mathematics*, Dover, New York NY, 1958.

(Smullyan, CP). Continuum Problem. In: Edwards (*EP*), pp. 207–212.

(Snell, *DMGPL*). Bruno Snell, *The Discovery of the Mind in Greek Philosophy and Literature* (reprint of *The Discovery of the Mind: The Greek Origins of European Thought*, tr. T. G. Rosenmeyer, 1953), Dover, New York, 1982.

(Sowa, *CS*). J. F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*, Reading, MA: Addison-Wesley, 1984.

(Stannett, 1990). Stannett, M., X-machines and the Halting Problem: Building a Super-Turing Machine, *Formal Aspects of Computing*, Vol. 2 (1990), pp. 331–341.

(Stone, *ToS*). I. F. Stone, *The Trial of Socrates*, Little, Brown and Company, Boston, 1988. The most plausible account I've read of how the conviction and execution of Socrates could have happened. It is particularly good at showing the impact of his theory of knowledge in other areas, such as ethics and politics.

(Stroyan & Luxemburg, *ITI*). K. D. Stroyan & W. A. U. Luxemburg, *Introduction to the Theory of Infinitesmals*, Academic Press, New York NY, 1976.

(Struik, *SBM*). D. J. (Ed.) Struik, *A Source Book in Mathematics, 1200–1800*, Harvard University Press, Cambridge MA, 1969.

(Styazhkin, *HML*). N. I. Styazhkin, *History of Mathematical Logic from Leibniz to Peano*, MIT Press, Cambridge MA, 1969.

(Tarski, CTFL). A. Tarski, "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, Vol. 1 (1936), pp. 261–405; English transl. "The Concept of Truth in Formalized Languages," in A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, tr. J. H. Woodger, Oxford Univ. Press, Oxford UK, 1956.

(Tarski, *IL*). Alfred Tarski, *Introduction to Logic and to the Methodology of Deductive Sciences*, second edition, revised, trans. Olaf Helmer, Oxford University Press, New York, 1946.

(Taylor, *PS*). A. E. Taylor, *Philosophical Studies*, Arno Press, New York NY, 1976; reprint of ed. published by Macmillan, London UK, 1934.

(Taylor, *VS*). A. E. Taylor, *Varia Socratica. First Series*, St. Andrew's University Publications, No. IX, James Parker and Co., Oxford, 1911; reprinted by Garland, New York, 1987.

(Thorndyke, *HMES*). Lynn Thorndyke, *A History of Magic and Experimental Science*, Columbia University Press, New York NY, 1923.

(Thornton & Thornton, *T&S*). Harry Thornton and Agathe Thornton, *Time and Style*, London, 1962.

(Thue, 1914). Axel Thue, "Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln," *Skrifter utgit av Videnskapsselskapet i Kristiania*, **I**. Matematisk-naturvidenskabelig klasse 1914, no. 10, 34 pp.

(Traugott & al., *OC*). E. C. Traugott, A. ter Meulen, J. S. Reilly, and C. A. Ferguson, *On Conditionals*, Cambridge University Press, Cambridge UK, 1986.

(Turing, CM&I). Alan M. Turing, Computing Machinery and Intelligence, *Mind*, Vol. 59, No. 236 (1950), pp. XXX–XXX; reprinted in Newman (*WM*, Vol. 4, pp. 2099–2123).

(Turing, OCN). Alan Turing, On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, series 2, vol. 42 (1936–7), pp. 230–265; corrections, ibid., vol. 43 (1937), pp. 544–546. Reprinted in Davis (*Undec.*, pp. 116–154).

(Untersteiner, *Soph.*). Mario Untersteiner, *The Sophists*, tr. Kathleen Freeman, Basil Blackwell, Oxford, 1954.

(van Heijenoort, LP). John van Heijenoort, "Logical Paradoxes," in *EP*, Vol. 5, pp. 45–51.

(Vickers, *ES*). Brian Vickers, *English Science, Bacon to Newton*, Cambridge University Press, Cambridge UK, 1987. Chapter 9 contains selections from Wilkins' *An Essay Towards a Real Character and a Philosophical Language.*

(von Mises, MPHU). Richard von Mises, Mathematical Postulates and Human Understanding. Reprinted in Newman (*WM*, pp. 1723–1754)

(von Neumann, *C&B*). John von Neumann, *The Computer and the Brain*, Yale University Press, New Haven, CN., 1958.

(Wang, *LWC*). W. S-Y. Wang, *Language, Writing, and the Computer*, W. H. Freeman, New York, 1986.

(Watson, 1913). John B. Watson, Psychology as the Behaviorist Views It, fill in

(Wescott, *SY*). W. Wynn Wescott (tr.), *Sepher Yetzirah or the Book of Creation*, Weiser, New York, NY, 1975.

(Westfall, RANC). Richard S. Westfall, The Role of Alchemy in Newton's Career. In: M. L. Righini & William R. Shea (eds.), *Reason, Experiment and Mysticism in the Scientific Revolution*, Science History Publication, New York NY, 1975, pp. 189–232.

(Weyl, *PM&NS*). Hermann Weyl, *Philosophy of Mathematics and Natural Science*, Princeton University Press, Princeton, 1949.

(Wheelwright, *Her.*). Philip Wheelwright, *Heraclitus*, Princeton University Press, Princeton, 1959.

(Wheelwright, *Pres.*). Philip Wheelwright (ed.), *The Presocratics*, The Odyssey Press, New York, 1966.

(Wiener, *DHI*). Philip P. Wiener, *Dictionary of the History of Ideas: Studies of Selected Pivotal Ideas* (five volumes), Charles Scribner's Sons, New York, 1973.

(Wightman, *SiRS*). W. P. D. Wightman, *Science in Renaissance Society*, Hutchinson University Library, London UK, 1972.

(Wilkins, *RC*). John Wilkins, *An Essay Towards a Real Character and a Philosophical Language*, Sa. Gellibrand & John Martin, London, 1668.

(Wittgenstein, ***PI*****).** Ludwig Wittgenstein, *Philosophical Investigations*, third edition, tr. G. E. M. Anscombe, Macmillan, 1953; Basil Blackwell & Mott, 1958.

(Wittgenstein, ***TLP*****).** Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, tr. by D. F. Pears & B. F. McGuinness, Routledge & Kegan Paul, London, 1961. Wittgenstein was dissatisfied with the first translation of the *Tractatus* (by C. K. Ogden & F. P. Ramsey), as he was, incidently, with Russell's introduction.

(Yates, ***AoM*****).** Frances A. Yates, *The Art of Memory*, University of Chicago Press, Chicago IL, 1966.

(Yates, ***GB&HT*****).** Frances A. Yates, *Giordano Bruno and the Hermetic Tradition*, University of Chicago Press, Chicago IL, 1964.

(Yates, Herm). Frances A. Yates, Hermeticism. In: Edwards (*EP*, Vol. 3, pp. 489–490)

(Yates, ***OPEA*****).** Frances A. Yates, *The Occult Philosophy in the Elizabethan Age*, Routledge & Kegan Paul, London, 1979.

(Yates, ***RE*****).** Frances A. Yates, *The Rosicrucian Enlightenment*, Routledge & Kegan Paul, London UK, 1972.

(Yonah & Shatzma, ***Enc.*****).** Michael Avi Yonah and Israel Shatzman, *Illustrated Encyclopedia of the Classical World*, Harper and Row, New York, 1975.

(Zeller, ***OHGP*****).** Eduard Zeller, *Outlines of the History of Greek Philosophy*, thirteenth edition, revised by Wilhelm Nestle, translated by L. R. Palmer, Humanities Press, New York, 1931.