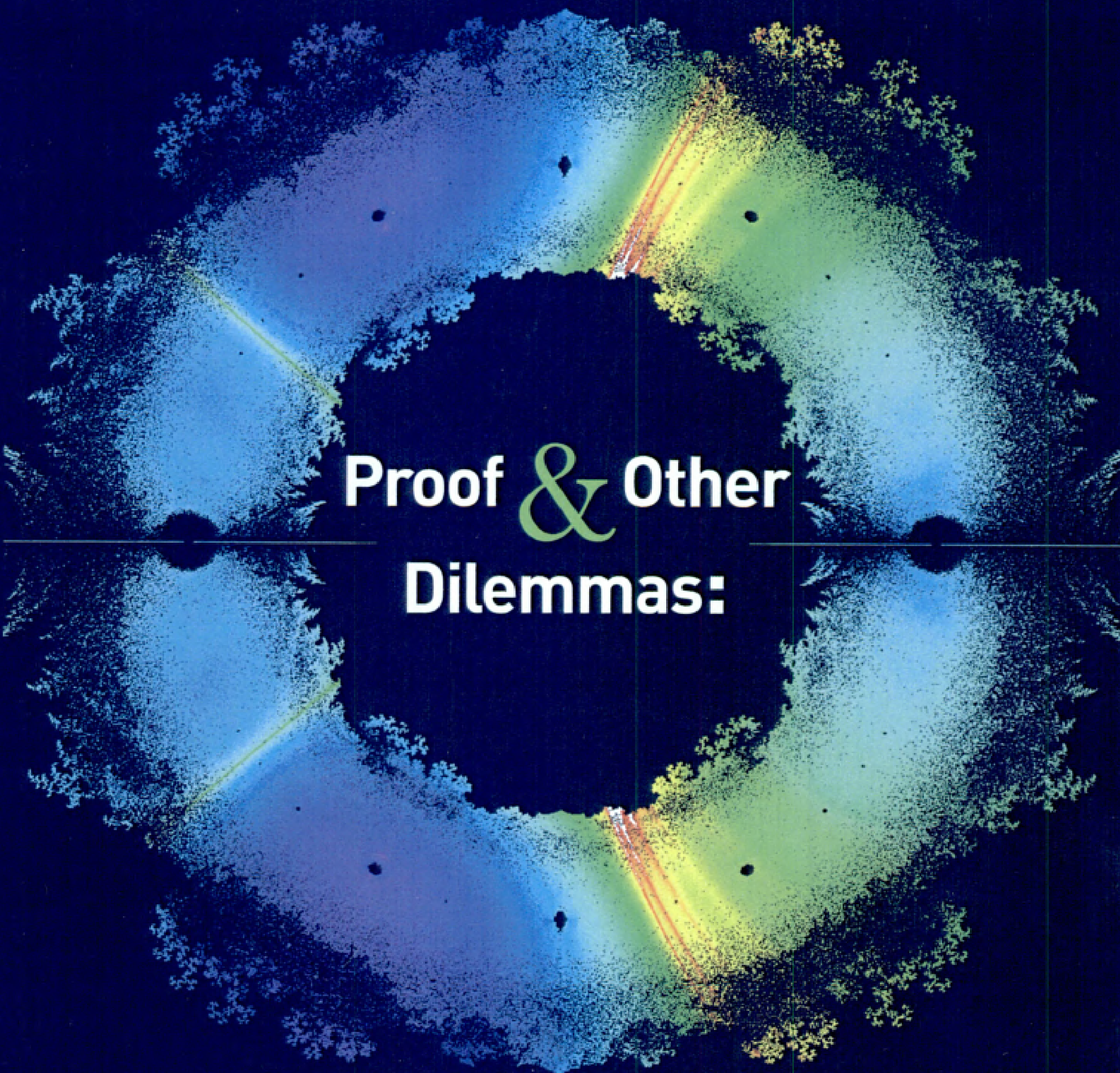
Mathematical Association of America

SPECTRUM



Proof & Other Dilemmas:

Mathematics and Philosophy

Bonnie Gold & Roger A. Simons, Editors

ISBN 978-0-88385-567-6

Library of Congress number 2008922718

Printed in the United States of America

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

Proof and other Dilemmas

Mathematics and Philosophy

edited by

Bonnie Gold and Roger A. Simons

Monmouth University *Rhode Island College*

Published and Distributed by

The Mathematical Association of America

Contents

Acknowledgements

We would like to express our appreciation to the Spectrum committee for their patience as we developed this book. We are especially grateful to J.D. Phillips and Sanford Segal, the two members of the committee who volunteered to read each chapter in its penultimate version and suggest changes that they felt would make each more accessible or historically accurate. We also appreciate Gerald Alexanderson's careful reading for misprints in the final version.

We are also grateful to all our authors for their patience as we pressed them multiple times to revise their chapters to make them accessible to a wider community. In addition, we especially would like to thank Professor Charles Chihara for reading and suggesting improvements in the first draft of the historical parts of the introduction.

We would like to thank Jonathan Borwein for the graphic found on the front cover of this volume.

More personally, we would like to thank our spouses, David Payne and Patricia Simons, for their support and patience through this project. What began as a sabbatical project of Bonnie Gold turned into a three-year project for both of us. We both spent many evenings and weekends on this project, reading and commenting on manuscripts, rather than with our respective spouses.

Finally, we appreciate the encouragement we have received during this project from numerous others and from a few of the authors.

Introduction

Bonnie Gold
Monmouth University

Section 1 of this introduction explains the rationale for this book. Section 2 discusses what we chose *not* to include, and why. Sections 3 and 4 contain a brief summary of historical background leading to contemporary perspectives in the philosophy of mathematics. Section 3 traces the history of the philosophy of mathematics through Kant, and Section 4 consists of an overview of the foundational schools. Section 5 is an annotated bibliography of sources for interesting recent work by some influential scholars who did not write chapters for this book. And finally, section 6 consists of very brief overviews of the chapters in this book.

1 The Purpose of This Book

This book provides a sampler of current topics in the philosophy of mathematics. It contains original articles by leading mathematicians, mathematics educators, and philosophers of mathematics written with a mathematical audience in mind. The chapters by philosophers have been edited carefully to minimize philosophical jargon, and summarize many years of work on these topics. They should thus provide a much gentler introduction to what philosophers have been discussing over the last 30 years than will be found in a typical book written by them for other philosophers. We have also included a glossary of the more common philosophical terms (such as epistemology, ontology, etc.). The chapters by mathematicians and mathematics educators raise and discuss questions not currently being considered by philosophers.

The philosophy of mathematics, starting about 1975, has been undergoing something of a renaissance among philosophers. Interest in foundational issues began receding and philosophers returned to more traditional philosophical problems. Meanwhile, some developments in mathematics, many related to the use of computers, have reawakened an interest in philosophical issues among mathematicians. Yet there is no book on these issues suitable for use in a course in the philosophy of mathematics for upper-level mathematics majors or mathematics graduate students, or for mathematicians interested in an introduction to this work. (Hersh's recent collection [Hersh 2005] contains many interesting articles related to the philosophy and sociology of mathematics, and is accessible to a similar audience, but it does not attempt, as we do, to cover the range of current discussion in the philosophy of mathematics.)

Our principal aim with this volume is to increase the level of interest among mathematicians in the philosophy of mathematics. Mathematicians who have been thinking about the philosophy of mathematics are likely to enjoy the variety of views in these papers presented in such an accessible form. Mathematicians who have never thought about philosophical issues but wonder what the major issues are should find several chapters to whet their interest. Those teaching courses in the philosophy of mathematics for upper-level mathematics undergraduates (or others with a similar mathematical background) should find it a useful collection of readings to supplement books on the foundational issues. Moreover, we hope to encourage more dialogue between two communities: mathematicians who are interested in the philosophy of mathematics, and philosophers who work in this field. We expect that most readers will not read every chapter in this book, but will find at least half to be interesting and worth reading.

2 What is not Included in This Book

A few words about our selection of topics for inclusion in this book are in order. We have not tried to include every topic that has ever been discussed in the philosophy of mathematics, or even everything currently being worked on. In part because we do not have adequate expertise to edit such articles, we have not included anything on the philosophy of statistics, which is currently a quite active field (although we do have a chapter on the philosophy of probability). More importantly, we have chosen *not* to include articles on the three foundational schools that developed in the late 19th and early 20th centuries: logicism, intuitionism, and formalism. They are described briefly later in this introduction, and much more thorough accounts of them appear in many books, including Stephan Körner's *The Philosophy of Mathematics*, Alexander George and Daniel Velleman's *Philosophies of Mathematics*, Marcus Giaquinto's *The Search for Certainty*, and Dennis Hesseling's *Gnomes in the Fog*. While there is still active work continuing in these fields, in our view the century from approximately 1865 to 1965 was an anomalous one for the philosophy of mathematics. What had seemed, prior to this period, to be the most certain form of human knowledge, mathematics, suddenly appeared to rest on shaky foundations. Thus essentially all work in the philosophy of mathematics during this period focused on trying to determine what basis we have for believing mathematical results. Gradually, problems were found with each of the foundationalist schools. Meanwhile new paradoxes did not appear despite an enormous growth in mathematics itself. As a result, the concern about mathematical coherence decreased, and philosophical attention began to return to more traditional philosophical questions. This book, then, concentrates on this new work, and complements the four books, just mentioned, that quite adequately discuss this foundational work.

Today there are many philosophers actively working in the philosophy of mathematics. A number of the better-known among them were invited to contribute to this book. Some of them declined due to prior writing commitments. However, several very well respected philosophers of mathematics *have* written chapters for this volume, and other viewpoints are well represented by some younger philosophers who were recommended by their mentors. Thus, most current viewpoints in philosophy are represented here. However, a single volume cannot hope to do this in full detail.

3 A Brief History of The Philosophy of Mathematics to About 1850

Although this book is concerned with recent developments in the philosophy of mathematics, it is important to set this work in the context of previous work. Thus I have written this historical section despite little expertise in the subject. Much of this material comes from, and is discussed in more detail in, chapter 1 of [Körner 1968]. Moreover, I am grateful to Charles Chihara for his many suggestions on how to improve my first version of this section. Any errors that remain here are my responsibility, not his.

The way a culture approaches mathematics and its use directly influences its philosophy of mathematics. Mathematics has been of interest to philosophers at least since ancient Greece. It has been used primarily as a touchstone to explore and test theories of knowledge. Traditionally, knowledge comes from two sources: sense perception and human reasoning. Mathematical knowledge has generally been taken as the archetypical example of the latter.

Plato is particularly important to any understanding of the history of the philosophy of mathematics, for two reasons. First, he is the earliest known philosopher who saw mathematics (which, for him, was synonymous with geometry) as central to his philosophical discussions. Ancient texts assert he viewed mathematics as so important that above the door of his Academy, Plato inscribed "Let no one who is not a geometer [or, "who cannot think geometrically"] enter." Plato used mathematical examples throughout his dialogues for various purposes. For example, in *Meno*, there is a famous sub-dialogue between Socrates and a slave boy. In it, Socrates leads the slave boy to discover that if you want to double the area of a square, you must take a square whose side is the diagonal of the original square. This discussion is used to explore an idea Plato wants to propose, of knowledge as memory from a previous life. There are several excellent books on Plato's philosophy of mathematics and the mathematics of Plato's time: for example, [Brumbaugh 1954] and [Fowler 1999].

Second, some of Plato's general philosophical views have resulted in his name being given to what is still seen, by philosophers today, as the default philosophy of mathematics, "platonism." That is, "platonism" is the view that (1) there are mathematical objects, (2) these are abstract objects, existing somewhere outside of space and time, (3) mathematical objects have always existed and are entirely independent of people, (4) mathematical objects do not interact with the physical world in any "causal" way—we cannot change them, nor can they change us—and yet, (5) we somehow are able to gain knowledge of them. These properties come from Plato's theory of "Forms," which appears in his later dialogues, primarily the *Republic* and *Parmenides*. Plato was struggling with our everyday world of appearance, trying to discern what is permanent and dependably true. This led him to the idea of the form of an object (say, a table) as a sort of ideal limit toward which objects of the physical world are striving but are imperfect copies. In this realm of forms live the assorted mathematical objects we work with: numbers, geometric objects, and so on. Objects in the realm of the forms are apprehended by reason, rather than by the senses. The appeal of viewing geometric objects, so central to Greek mathematics, this way is apparent. We see imperfect lines and points in the physical world and can easily imagine a perfect point and line. Mathematical statements are necessarily true, because they describe objects in this unchangeable realm. Objects in the physical world "participate in" the forms that describe them, and, because they are only imperfect likenesses, are only approximately described by mathematical theorems.

Aristotle objected to abstracting properties of objects into an independent existence. Rather, you can discuss these abstracted *properties*, but they reside in the objects they're abstracted from. Mathematical statements are then idealizations of statements about objects in the physical world. To the extent that these idealizations are accurate representations of the physical objects they're abstracted from, mathematical theorems can be approximately applied to physical objects. Two other contributions Aristotle made to the philosophy of mathematics were a discussion of infinity, and the beginnings of logic. Aristotle's distinction between potential infinities (basically, what happens when we take the limit as $x \to \infty$) and actual infinities (such as the set of integers, real numbers, etc.) was important historically in mathematicians' hesitation to accept many developments involving actual infinities.

Gottfried Wilhelm Leibniz was, of course, one of the founders of calculus, but he also made a substantial contribution to logic and to philosophy. He believed that by developing a systematic calculational logic (a "calculus ratiocinator"), one could represent much human reasoning and resolve many differences of opinion. He began to develop such a system, and introduced many of the modern logical concepts: conjunction, disjunction, negation, etc. (None of this, however, was published during his lifetime.) For Leibniz, mathematical facts are truths of reason, "necessary" truths whose denial is impossible (as opposed to truths of fact, that are "contingent," that just happen to be true in this world, and whose denial is possible). Mathematical facts are true in "all possible worlds" (a terminology he introduced).

John Stuart Mill was a complete empiricist about mathematics as about everything else. He believed that mathematical concepts are derived from experience and that mathematical truths are really inductive generalizations from experience. There are no necessary truths. Thus every mathematical theorem can, in principle, be found to be false and in need of revision. Mathematical truths are about ordinary physical objects. Geometrical propositions are inductively derived from our experience with space, and are taken to mean that, the more closely physical objects approach these idealized geometrical objects, the more accurately the theorems can be applied to them. Statements such as "$2 + 3 = 5$" is a generalization about how many objects you get when you put together a pile of two and a pile of three physical objects. However, for people, such inductive generalizations are a psychological necessity, because they come from very deep and invariant experiences. These experiences create an appearance of mathematical facts being necessarily true.

For **Immanuel Kant**, mathematics provided central examples for his classification of knowledge. Knowledge of propositions was classified into *a priori* or *a posteriori*. Meanwhile, propositions were classified as synthetic or analytic. A proposition is known *a priori* ("from the former"—before experience) if it is known without any particular experiences, simply by thinking about it. A proposition is known *a posteriori* if knowledge of it is gained from experience or via the senses. An "analytic" proposition is one whose predicate is contained in its subject. For example, "all squares are rectangles" is analytic because the definition of square (as "a rectangle with congruent sides") contains the requirement that it be a rectangle. The canonical example of an analytic proposition is "all bachelors are unmarried." Many mathematical and logical truths are analytic and are known *a priori*, as with "all squares are rectangles." A proposition that is not analytic is "synthetic." According to Kant, most truths about the world—"Mount Everest is the highest mountain in the world," for example—are synthetic, and are known *a posteriori*. It is generally believed that no analytic propositions can be known *a posteriori* (although a modern philosopher, Saul Kripke, has disputed this).

This leaves the category of synthetic propositions that are known *a priori*. According to Kant, our intuitions of time and space, which give us facts about the real numbers ($\mathbb{R}^3$ in particular) and the integers (such as $2 + 5 = 7$), are synthetic, yet are known *a priori*. We do not get them simply by analyzing their definitions, but rather by thinking about space and time. (Frege disagreed with this, at least in the case of arithmetic facts; he viewed them as analytic, and this was part of the point of his *Foundations of Arithmetic*.) Nonetheless, actual experience with space or time is not required to get this knowledge. In Kant's case, by space, he meant Euclidean space; that is, Euclidean geometry gives us our intuition of space. Thus Euclidean geometry is the inevitable necessity of thought, rather than being of empirical origin. The integers come from our intuition of time in the form of one moment, then the next moment, and so on. (This idea first appeared much earlier, in Plato's *Timaeus* 39b-c.) Another distinction Kant makes is between concepts we can both perceive and construct—such as the concept of two objects—those we can construct but not perceive—such as $10^{10^{10^{10}}}$—and those we can neither construct nor perceive, but are simply "ideas of reason" because they are consistent—such as actual infinities.

4 The Foundational Problems and the Three Foundational Schools

In the nineteenth century, three events occurred that caused both mathematicians and philosophers to reassess their views of issues such as what mathematics is about, how we acquire mathematical knowledge, and how mathematics can be applied to the physical world.

First came the inconsistencies in the use of limits in calculus. Soon after the introduction of calculus, there were concerns about foundational issues. Derivatives were found by taking a ratio of two infinitesimal quantities and then treating the denominator as if it were zero (even though if the denominator *is* zero, one could not even have formed the quotient). Bishop George Berkeley, in *The Analyst*, 1734, inveighed against "ghosts of departed quantities." Maclaurin responded by showing how one can derive calculus results via contradiction in the "manner of the ancients," the method of exhaustion—thus, calculus is simply a short-cut to legitimate results. Lagrange (1797) responded to the problem by trying to use power series to get rid of limits. This introduced its own problems, in the absence of some way of seriously considering issues such as divergence. For example, Grandi, in 1703 (see [Burton 2003], p. 567) set $x = 1$ in

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots \quad \text{to get} \quad \frac{1}{2} = 0.$$

These difficulties were of increasing concern to mathematicians in the 1800s, and the need for developing textbooks for university students eventually led Cauchy, in the 1820s, to give careful definitions of continuity, differentiability, and the integral. But Cauchy did not notice the need for a distinction between pointwise and uniform convergence of functions. As a result, he stated a false theorem about convergence of sums of continuous functions. This led to the development of careful definitions of the limits by Weierstrass and of the real numbers by Dedekind. Fourier series introduced a new set of complications: when did they converge to a function, and what exactly was a function anyway? More broadly, mathematicians began to be concerned about the foundations of mathematical beliefs—how can we be sure that what we develop is free of contradictions? This concern led to logicism, the attempt to reduce all of mathematics to logic. This work began with Gottlob Frege's *Foundations of Arithmetic* in 1884. In his 1879 *Begriffsschrift*, he developed the first fully formalized axiomatic development of the sentential calculus; he also introduced quantifiers and expanded his calculus to a predicate calculus. This was the basis for modern predicate logic, a major advance over the Aristotelian logic that had dominated for centuries. It provided him the language to formalize arithmetic. It was hoped that if there were any contradictions in mathematics, they would inevitably be found before they could cause any damage once everything was reduced to logic. One thing this formalization of arithmetic accomplished was to make statements such as $2 + 5 = 7$ a consequence of the definitions of the numbers involved, and thus turn them into analytic propositions.

The second event was the development of non-Euclidean geometries: Lobachevsky (lectured on in 1826, published about 1835), Bolyai (published in 1831 as an appendix to a book by his father), Gauss (who apparently discovered it earlier but did not publish it), and later Riemann (1854). Apparently because of the obscurity of the journals in which the work of Lobachevsky and Bolyai published, it was not until Riemann's work that the world-view of Kant was finally rejected. (Kant's world-view (see section 3, above) was that Euclidean space is the inevitable necessity of thought, rather than being of empirical origin.) Euclidean geometry was no longer the science of space—it is still far from clear which geometry is best to describe actual physical space. This revolutionary development threw mathematics out of the physical world (though, of course, not out of its usefulness in describing that world). It also led to the use of axiomatics as a way of discovering new mathematics.

The third event causing a revolution in the philosophy of mathematics was the discovery of contradictions (paradoxes) in naïve set theory. These contradictions were discovered not only for the work of Cantor (which many mathematicians were already suspicious of, as it dealt with "completed infinities") but right there in the careful work of Frege. Frege's *Grundgesetze der Arithmetik* (Fundamental Laws of Arithmetic) is a work of logicism reducing the truths of arithmetic to theorems of logic. The second volume was at the publisher in 1903 when Russell wrote to Frege, informing him of the inconsistency of his system via the Russell paradox. (The set consisting of all sets that are not members of themselves both *must be* and *must not be* a member of itself. That is, let $A = \{B : B \text{ is a set and } B \notin B\}$. Then both $A \in A$ and $A \notin A$ lead to contradictions.) Thus, the elementary step of forming the set of all objects having a given property can lead to a contradiction. Since mathematicians frequently form sets this way, this discovery shook a larger portion of the mathematical world than the others. Instead of the occasional misuses of limits, which were viewed as the result of bad mathematical taste, the view now was that there was a *crisis in the foundations of mathematics*. Was all of mathematics a house built on shifting sands? In response to this crisis, two additional foundational schools, intuitionism and formalism, were developed. Logicism was also further developed, by Russell and Whitehead, Zermelo, and others, trying to mend the problems in Frege's account.

Logicism is the thesis that mathematics is a sub-branch of logic, that all theorems of mathematics can be reduced to theorems of logic. Logic had experienced significant development in the nineteenth century in the work of Boole, De Morgan, C. S. Peirce, and Venn, among others, as well as, of course, Frege. This work made logic a very systematic study of correct rules for reasoning. Therefore, it seemed plausible that if all of mathematics could be deduced from logic, mathematics would be free of contradictions and its foundations firm. In addition, the work of Peano giving axioms for the natural numbers, of Dedekind building the real numbers, and of Weierstrass defining limits, gave logicists much material needed to reduce mathematics to logic. Frege began this work with his *Grundlagen der Arithmetik* (Foundations of Arithmetic) and continued it with his two-volume *Grundgesetze der Arithmetik* (Fundamental Laws of Arithmetic). However, the Russell paradox meant that a different approach needed to be taken. Russell and Whitehead developed one such approach in *Principia Mathematica*, an enormous work comprising three volumes and over 2000 pages. Their hope was to show in this work that all of mathematics (or at least, number theory) could be reduced to logic. Russell had analyzed the Russell paradox and other paradoxes of set theory, and determined that all of them involved defining a set by using a larger set of which the set being defined was a member, which he called a "vicious circle." He believed that, as long as one avoided using vicious-circle definitions (also called "impredicative definitions"), one could avoid paradoxes. To do this, the set theory developed in *Principia Mathematica* builds sets in a hierarchy, a type-theory, with sets of the lowest type being individuals. On the next level are sets composed of these individuals. At each level, sets are built up of members that are sets from previous levels. The known large sets that lead to contradictions cannot be constructed in this system. (For brevity, the description here is significantly simplified. Their actual approach used propositional functions, rather than sets, as the basic objects on which everything else was built, and a "ramified" theory of types. See *Ontology and the Vicious-Circle Principle* [Chihara 1973], chapter 1, for a good description.)

Logicism in this revised form had three significant problems, which largely led mathematicians to lose interest in it. First, with a rigid type-theory, many important mathematical theorems not only cannot be proven, they cannot even be stated. For example, the least upper bound of a bounded set of real numbers is defined in terms of the set of real numbers. Therefore it must be of a higher type than the real numbers. Thus, this least upper bound cannot be, in this type-theory, a real number. To overcome this problem, an additional axiom was added to their system, called the axiom of reducibility. This axiom essentially says that a set that is defined at a higher level using only sets at some lower level is equivalent to some set that appears at the first level above all those involved in its definition. The problem with this axiom is that there is no justification for it within logic (and there are some concerns that it might allow the paradoxes to reappear). Hence, the program of reducing mathematics to logic fails: either you cannot get well-known theorems, or you must add a principle that is not purely logical.

There is a similar problem with the axiom of infinity. For much of mathematics, we need infinite sets. Yet their existence simply does not follow from other axioms. Russell and Whitehead introduced it as an axiom, but cannot justify it based purely on logic. Later logicians have attempted to overcome these two problems, most notably Quine, but no one has managed to build up all of mathematics purely from logical principles.

Third, Gödel's incompleteness theorem dealt a very significant blow to even the *possibility* of deriving all of mathematics from logic. At least for consistent first-order, recursive axiomatizations of number theory, this theorem says that if they are sufficiently strong to prove normal arithmetic properties, then there are theorems that are true but not provable in such systems. Hence, one simply cannot get all of mathematics from (at least first-order) logic.

Intuitionism is the thesis that mathematical knowledge comes from constructing mathematical objects within human intuition. Intuitionism's ancestors were Kant, Kronecker, and Poincaré. Kant contributed an intuition of the integers from our *a priori* intuition of time. Kronecker was famous for his statement "God made the natural numbers; all else is the work of man." He objected to any mathematical object that could not be constructed in a finite way. In particular, he fought Cantor's transfinite numbers. Poincaré viewed logic as sterile, and set theory as a disease. On the other hand, he viewed mathematical induction as a pure intuition of mathematical reasoning.

L.E.J. Brouwer, the founder of intuitionism, believed that the contradictions of set theory came from inappropriate dependence on formal properties, including logic. In particular, the use of the law of the excluded middle (that either a statement P or its negation ~P must be true) with completed infinities or with proofs of existence is illegitimate and dangerous to the coherence of mathematics. Brouwer started with Kant's idea that our intuition of time is the basis for the natural numbers. Mathematical objects are mental constructions, which Brouwer described as "intuited non-perceptual objects and constructions which are introspectively self-evident." ([Körner 1968], p. 120) Completed infinities cannot be inspected or introspected, and so are not part of mathematics. A mathematical statement is true "only when a certain self-evident construction had been effected in a finite number of steps." ([Burton 2003], p. 661) To prove a proposition of the form "P or Q," one needs to prove P or to prove Q. To prove that $(\exists x)P(x)$, one needs to give a construction of an object and show that it satisfies P.

The rejection of completed infinities causes problems in the construction of real numbers. To define a real number, the intuitionist must, for example, give an algorithm that produces a sequence of rational numbers and give a proof that that sequence converges.

Many standard theorems are not intuitionistically true. For example, the standard proof of the Intermediate Value Theorem involves repeatedly bisecting an interval on which the function changes from being below the desired value C to being above it (or vice versa), maintaining that property in the sub-interval chosen. However, intuitionistically, one cannot always *determine* whether a given real number is greater than, equal to, or less than C. For example, let

$$a_n = \begin{cases} 1 & \text{if } 2n \text{ is the first even integer that is not the sum of two primes, } n > 1,\ n \text{ even} \\ -1 & \text{if } 2n \text{ is the first even integer that is not the sum of two primes, } n > 1,\ n \text{ odd} \\ 0 & \text{otherwise} \end{cases}$$

Define the real number $r = \sum_{n=2}^{\infty} a_n 10^{-n}$. Both intuitionistically and classically, r is a well-defined real number: to calculate its nth digit, just check if all even integers from 4 to $2n$ can be written as the sum of two primes. If the Goldbach conjecture is true, $r = 0$. If it is false and first fails at a multiple of 4 (i.e., n as used above is even), $r > 0$. If it first fails at an integer congruent to 2 modulo 4, $r < 0$. You can calculate r to whatever degree of accuracy you wish, simply by trying to decompose the appropriate values of n into sums of two primes. But while, classically, r must be either positive, negative, or zero, intuitionistically it is none of these until we decide the Goldbach conjecture. One can easily use r to give a function that shows that the Intermediate Value Theorem is not true intuitionistically, not just that there is a problem with the usual proof.

Intuitionism was developed in the same period that many abstract areas of modern mathematics—topology, functional analysis, etc.—were being developed. Most mathematicians were more interested in exploring these new developments than in retreating inside the shell of intuitionism.

There are many philosophical problems with intuitionism as well. If mathematical objects are mental constructions, there is no good reason to believe that two people will construct the same objects or have the same theorems. It is also not clear why mathematics is so useful in the world. Furthermore, much of modern physics uses mathematical objects (from functional analysis, for example) that intuitionists do not accept.

Intuitionism initially received enthusiastic support from Hermann Weyl (although he fell away from it later). Arend Heyting extended Brouwer's work in intuitionism and made Brouwer's often mystical and obscure writing much more accessible. However, because so many theorems of standard mathematics cannot be proven intuitionistically, very few mathematicians were inclined to adopt intuitionism. It required giving up too much mathematics just to avoid a few contradictions with extremely huge "sets." In the 1960s, Errett Bishop developed a variation on intuitionism, which he called constructivism (see [Bishop 1967]). He developed many theorems that are, using classical logic, equivalent to standard theorems but are constructively true. Thus, at least in analysis, one needs to give up less mathematics to be a constructivist than to be an intuitionist. This led to some renewed interest in the subject, but still has not led very many mathematicians to abandon classical mathematics.

Formalism is less well-defined. It is not clear that many serious mathematicians ever asserted the most extreme version of what is called formalism, that mathematics is just a formal game.[1] This view of mathematics is extremely unhelpful philosophically: it does not explain why we choose the axioms we choose, why mathematics is applicable to the world, why anyone would bother studying mathematics at all.

This extreme characterization of formalism appears to come from combining two parts of Hilbert's work. In his Foundations of Geometry (*Grundlagen der Geometrie*), he fixes some incompletenesses in Euclidean geometry, adopting an axiom system based on three undefined objects—points, lines, and planes—and three undefined relations—incidence (a point lying on a line), order (betweenness), and congruence. He makes it clear that, while the intuition behind the axioms comes from what we call points, lines, and planes, they could just as well stand for any objects—say, tables, chairs, beer mugs—as long as those objects satisfy the axioms. This work of Hilbert is one of the early works of modern mathematics, where, instead of working entirely within one mathematical structure, one sets up definitions and axioms and then proves theorems about the whole class of objects that satisfy the definitions and axioms.

Hilbert's proof of the consistency of his axioms for geometry reduces the question of the consistency of those axioms to the consistency of arithmetic. This brings us to the second part of Hilbert's work that is relevant for formalism. This is the "Hilbert program," aimed at restoring confidence in mathematics after the contradictions, described above, that came from work in the foundations of analysis and from naïve set theory. In part, his program was a reaction to what he considered the pernicious affect that intuitionism was having on mathematicians. He was determined to put mathematics on a sound footing without giving up large parts of mathematics in the process. The program is to first set up each field of mathematics as a formal theory, consisting of undefined terms and axioms. A proof in such a theory is a finite sequence of formulas, each of which is either an axiom or follows from earlier formulas by finitary logical rules of inference. One then investigates several metamathematical questions about the systems thus developed.

First, is the theory consistent? This can be investigated in one of two ways. One is to give a model of the theory. Usually this involves picking an already known mathematical structure (such as the integers). Then one interprets each of the undefined terms of the theory as objects within that structure in such a way that all of the axioms can be shown to be true theorems about the structure. When the structure involved is infinite, this then reduces the consistency of the original theory to the question of whether the axioms for the structure used to interpret the theory are consistent. Thus it is called a "relative consistency proof." Hilbert (in his *Grundlagen der Geometrie*) had given such proofs for Euclidean and non-Euclidean geometry by interpreting them within the real algebraic numbers. Thus, as long as the arithmetic of the real numbers is consistent, so is both Euclidean and non-Euclidean geometry. But this kind of consistency proof

---

[1] An exception is apparently von Neumann, who allegedly said "We must regard classical mathematics as a combinatorial game played with the primitive symbols . . . " [von Neumann 1966, pp. 50–51]. There is a quotation floating around, attributed to Hilbert: "Mathematics is a game played according to certain simple rules with meaningless marks on paper." This quotation appears for the first time in E.T. Bell, without citation—it may well have been made up by Bell. In fact, Hilbert, in [Hilbert 1919, p. 19], said "Mathematics is **not** like a game in which the problems are determined by arbitrarily invented rules. Rather, it is a conceptual system of inner necessity that can only be what it is and not otherwise." (translated by Michael Detlefsen, emphasis mine).

does not rule out the possibility that all of the theories involved are inconsistent. In addition, it is not using strictly finitary reasoning, and thus does not provide the foundation that is needed.

The second way consistency can be investigated is to show, in a finitary way, that it is not possible to derive a contradiction (for example, the statement $0 = 1$) from the axioms. This would then be an *absolute* consistency proof. It would not depend on another system (except, of course, the logic involved, which is finitary and might be acceptable to intuitionists). Of course, if one could give this kind of consistency proof for arithmetic, it would provide an absolute proof of the consistency of geometry, since a relative consistency proof had reduced the consistency of geometry to that of arithmetic.

Second, is the theory complete? This has a syntactic and a semantic meaning. Semantically, can all truths about the structure involved be proven from the axioms? If an axiomatization is not complete, then it has not captured all the relevant features of the mathematical structure it is axiomatizing, and there is a need to find further axioms so as to fully represent the structure. Syntactically, if an axiomatization is complete, every sentence or its negation is derivable from the axioms (since every sentence is either true or its negation is true in the structure).

Third, are the axioms independent of each other, or can some be eliminated? One usually shows independence by giving a structure in which all but one of the axioms are true, and the remaining one fails. This is the least important question, more an aesthetic issue than one central to the adequacy of the theory. But as mathematicians tend to like clean results, it is preferable to find axioms that are independent. Hilbert, in his *Grundlagen der Geometrie*, showed that many of his axioms were independent, though, given the tediousness of going through all combinations, he did not show that all were.

The foundational school called formalism contains as its core the view that to set mathematics on firm foundations, one should investigate these questions for the various structures and theories that make up mathematics. This led to the development of the field called proof theory, which investigates these metamathematical questions.

Unfortunately for Hilbert's program, two results of Gödel showed that the program could not work. His first incompleteness theorem showed that any consistent first-order axiomatization for the natural numbers that can be described recursively (basically, in a finitist way), and that is sufficiently strong to prove most of the standard theorems of number theory, is incomplete. That is, there are truths about arithmetic that cannot be proven within that axiomatization. (Actually, the result Gödel proved required a little more, called ω-consistency; the result was improved by Barclay Rosser to simply require standard consistency.) Thus, one cannot capture all truths about the integers within a finitistic system. His second incompleteness theorem was even more devastating. Given any consistent, recursive system of (first-order) axioms that is sufficiently strong to do a significant amount of mathematics[2], it is impossible to prove the consistency of the system within that system. Thus, there is no point in looking for a finitary proof of consistency. There has been continuing work in proof theory investigating properties of axiom systems, but there does not appear to be any hope of reviving Hilbert's original program. Gödel proved a third important theorem relevant to the Hilbert program, the completeness theorem for first-order

[2] Here, "sufficiently strong" represents a technical requirement involving being able to represent the primitive recursive functions within it and derive some standard number theoretical results; for details, see any standard textbook on mathematical logic.

logic. This says that every consistent set of first-order statements has a model. That is, our system of first-order logic is complete: in it, every first-order statement which is true in every model can be proved. Thus, semantic consistency (having a model), for first-order theories, is equivalent to syntactic consistency (not being able to derive a contradiction). Second-order theories, however, may be consistent without having any models.

Of these foundational schools, only logicism can really be called a philosophy of mathematics, as the other two do not really provide answers to all of the traditional philosophical questions: "what is the nature of mathematical objects," "what is the nature of mathematical knowledge," and "how can mathematical results help us understand physical world?" Intuitionism does not answer the last; formalism does not answer the first or the third (and, because of Gödel's results, does not answer the second either). Logicism's answer to all of these questions reduces to the similar questions about logic. However, since there are serious problems in reducing mathematics to logic, logicism does not settle these questions either. But for the first three-quarters of the twentieth century, work on foundations replaced almost all other discussion about the philosophy of mathematics.

More detailed discussions of the three foundational schools can be found in [Burton 2003]; [Körner 1968] and [George/Velleman 2002] are books, aimed at the same audience as this book, devoted to a thorough discussion of these views. Also, [Giaquinto 2002] is an accessible book that gives a good discussion of what work has been done in each of these schools.

4.1 Other Philosophers in This Period

There are two philosophers who wrote a substantial amount about mathematics during this period, but were not part of any of these foundational schools. One was **Edmund Husserl**, who developed phenomenology. He had a Ph.D. in mathematics, and his *habilitation* dissertation was *On the Concept of Number* (1887), which was later expanded to *Philosophy of Arithmetic*, published in 1891. This book attempted a psychological foundation of arithmetic, and preceded his phenomenological work, which was first published in 1900 in *Logical Investigations*. Husserl also has a very fine (and influential) essay, called "The Origin of Geometry," that usually appears as an appendix to his *The Crisis of European Sciences and Transcendental Phenomenology*. Derrida's Ph.D. thesis is a response to it. Husserl is quite difficult to read. Richard Tieszen has worked on making Husserl accessible, as well as answering philosophical objections to Husserl's work; see his *Phenomenology, Logic, and the Philosophy of Mathematics* [Tieszen 2005].

Ludwig Wittgenstein is another influential philosopher of this period who is also not easy to read. His work focuses on "language games," or the relation between language, as we use it, and reality. His initial work on this topic in the *Tractatus Logico-Philosophicus* (1922) set the stage for his work on the philosophy of mathematics in *Philosophical Remarks* (1929–30), *Philosophical Grammar* (1931–33), and later in *Remarks on the Foundations of Mathematics* (1937–44). According to the Stanford Encyclopedia of Philosophy, Wittgenstein maintains that mathematical propositions differ from real propositions. Mathematical statements do not refer to anything real, but their content comes from their syntax. "On Wittgenstein's view, we invent mathematical calculi and we expand mathematics by calculation and proof, and though we learn from a proof that a theorem *can* be derived from axioms by means of certain rules in a particular way, it is *not* the case that this proof-path pre-exists our construction of it." (http//plato.stanford.edu/entries/wittgenstein-mathematics/) He views mathematics as a human invention, and no mathematics exists until we discover it. Wittgenstein is thus a precursor of some social-constructivist views of mathematics.

5 More Recent Work That is Worth Reading but is Not Represented Here

As I mentioned in the first section of this introduction, this book consists of original articles by philosophers, mathematicians, and mathematics educators, most summarizing work over a period of years. To put this book together, I invited people whose work I had read and admired to write a chapter for this volume. I got a relatively good response, and thus this volume covers a fairly wide range of contemporary issues. However, in part because I was often asking very senior people in the field, there were a number of excellent writers on the philosophy of mathematics who declined to participate in this project. You'll certainly find suggestions for continued reading on any of the topics in this book in the bibliographies of the individual chapters. However, I want to take some space here to recommend some other very good places to learn more about the philosophy of mathematics. Full bibliographic references for these books and articles are in the Bibliography at the end of this introduction.

5.1 Logicians with a Philosophical Bent

Two logicians who have done a significant amount of very thoughtful and careful work in the philosophy of mathematics have recently collected that work in books: **Solomon Feferman**'s (math.stanford.edu/~feferman/) *In the Light of Logic* [1998] and **William Tait**'s (home.uchicago.edu/~wwtx/) *The Provenance of Pure Reason: Essays in the Philosophy of Mathematics and Its History* [2005]. I recommend both books highly.

5.2 Philosophers

There are many philosophers working in the philosophy of mathematics, almost all of them working on questions of the nature of mathematical objects and of mathematical knowledge: the debate, represented and summarized in this volume by the chapters by Balaguer, Chihara, Linnebo, and Shapiro, of platonism versus nominalism. All of the philosophers listed below have written a lot more than is mentioned here, of course; but I'm pointing to those I think are likely to be interesting to mathematicians.

Paul Benacerraf, at Princeton (philosophy.princeton.edu/components/com_faculty/documents/paulbena_cv.pdf) wrote two seminal papers that initiated the move in the 1970s, by philosophers of mathematics, back to traditional philosophical questions and away from foundations: "What Numbers Could Not Be" [1965] and "Mathematical Truth" [1973]. They are still well worth reading. He also, with Hilary Putnam, edited a book of readings in the philosophy of mathematics. It has gone through two editions, with a quite different selection of papers in the second edition. Both editions are worth looking at. His two articles mentioned above are reprinted in the second edition.

John Burgess (www.princeton.edu/~jburgess/), also at Princeton, works in logic (philosophical and mathematical) and the philosophy of mathematics. What he says about mathematics is very careful and correct. However, since he works in the philosophy department, his interests have been turning more and more toward technical philosophical issues. His "Why I Am Not a Nominalist" [1983] is quite accessible. I have not read his two recent books: *A Subject with No Object* [Burgess/Rosen 1997] and *Fixing Frege* [2005]. I looked at the former and decided that it was far more technical than I could handle without devoting months to it. I do hope to look at the latter once this book is finished.

Imre Lakatos combined the approaches of philosophers of science Thomas Kuhn and Karl Popper and applied it to mathematics. His best-known work is *Proofs and Refutations* [1976], a lively dialogue about Euler's theorem that $v - e + f = 2$ (where v represents the number of vertices, e the number of edges, and f the number of faces) for a polyhedron. It shows how cases that are counterexamples motivate revisions of the hypotheses of the theorem and the definition of polyhedron. This provides, according to him, an example of how mathematics develops.

Penelope Maddy (www.lps.uci.edu/home/fac-staff/faculty/maddy/), at the University of California at Irvine, started her career as a student of Burgess and a platonist. Her first book, *Realism in Mathematics* [1990], described an unusual form of platonism in which mathematical objects are located in the physical world. This view was broadly attacked by other philosophers. Her current direction, a naturalist approach to the philosophy of mathematics (*Naturalism in Mathematics* [1997]), is one that takes science as the standard by which all knowledge is to be judged. Knowledge of anything, including mathematics, must be justifiable through our best scientific theories, in particular, empirical psychology, linguistics, etc.

Charles Parsons (www.fas.harvard.edu/~phildept/parsons.html), at Harvard, works in the philosophy of mathematics as well as in logic and in other fields of philosophy. His article "Mathematical Intuition" [1979–80] is a fairly interesting discussion of how one can have intuitions of mathematical objects such as numbers and sets. However, it will be disappointing if you are expecting something like what Poincaré wrote on the topic. Many of his papers in the philosophy of mathematics are collected in [Parsons 1983].

Hilary Putnam (www.fas.harvard.edu/~phildept/putnam.html), also at Harvard, works on philosophy of mathematics, philosophy of science, and other fields of philosophy. His article, "Mathematics without Foundations" [1967], is one of the early articles moving philosophy of mathematics back from foundations to more traditional philosophical problems. Some of his work in the philosophy of mathematics is collected in [Putnam 1985].

Michael Resnik (http://philosophy.unc.edu/resnik.htm), at the University of North Carolina, Chapel Hill, is a structuralist (of a slightly different sort than Stewart Shapiro, who has an article in this volume). His book, *Mathematics as a Science of Patterns* [1997], is quite readable once you are used to philosophical terminology. Given that many mathematicians assert that mathematics is the science of patterns, the book is worth reading to see how philosophers establish such an assertion.

5.3 People Working in the History and Philosophy of Mathematics

Several people work on the boundary between philosophy of mathematics and history of mathematics. One is **Kenneth Manders** (www.pitt.edu/~philosop/people/manders.html), at the University of Pittsburgh. He is primarily a philosopher (and logician) with a strong mathematical background, but his arguments are very carefully historically based. Unfortunately, he rarely publishes. One published article is "Domain extension and the philosophy of mathematics"

[1989]. I have a very interesting preprint, "Why Apply Math?" from 1999, and another, "Euclid or Descartes: Representation and Responsiveness." Both are very carefully and thoughtfully written, but the only way to get them is to write to him.

Another is **Leo Corry** (http://www.tau.ac.il/~corry/), at Tel Aviv University. He is more of a historian of mathematics, but he asks philosophical questions. For example, his *Modern Algebra and the Rise of Mathematical Structures* [2004] investigates how the notions of what was meant by "algebra" and "mathematical structure" developed over the last two centuries.

A third is **Paolo Mancosu** (http://philosophy.berkeley.edu/mancosu/); see his "On Mathematical Explanation" [2000]. In addition to being interesting in itself, it mentions several other articles on this topic.

A fourth is **Howard Stein** (https://philosophy-data.uchicago.edu/index-faculty.cfm#Stein), at the University of Chicago, who works on the history and philosophy of mathematics and physics. Three articles worth looking at are "Yes, but . . . : Some Skeptical Reflections on Realism and Anti-realism" [1989], "Eudoxos and Dedekind: On the Ancient Greek Theory of Ratios and its Relation to Modern Mathematics" [1990] and (do not be put off by the title) "Logos, Logic, and Logistiké: Some Philosophical Remarks on 19th Century transformation of Mathematics" [1988].

5.4 People Working in Mathematics Education

By the very act of teaching mathematics, one takes a position on how people acquire mathematical knowledge, which has both psychological and philosophical aspects. Therefore many people whose research is in mathematics education have interesting philosophies of mathematics. I contacted several of them, and, as it turns out, the person whose work I find most interesting *has* made a contribution to this book, but several others whose work I also respect were either unwilling or unable to do so.

Ed Dubinsky (http://www.math.kent.edu/~edd/) has worked applying an interpretation of Piaget's work to higher-level mathematics education. He is very active (though now retired), and has gathered a large community of mathematics educators who work with him. He started the Research in Undergraduate Mathematics Education Community, which later branched into the SIGMAA on RUME. He started out as a functional analyst. In his early work in mathematics education, he used a computer-based language, ISETL, to help students understand abstract mathematical objects. For example, see his "Teaching mathematical induction, I/II" ([1986], [1989]). The basic theory that he developed, APOS theory (standing for Action, Process, Object, Schema), describes how students gradually develop more sophisticated concepts through a process of reflective abstraction. A description can be found in "A theory and practice of learning college mathematics" [1994]. It is applied to student understanding of functions in "Development of the process conception of function" [1992], and to abstract algebra in "Development of students' understanding of cosets, normality and quotient groups" [1997]. An overall framework is given in "A framework for research and curriculum development in undergraduate mathematics education" [1996]. Dubinsky views his philosophy as inseparable from his educational theory and practice, and declined to write an article for this volume because he felt that his work already expresses his philosophical position adequately.

Paul Ernest (http://www.people.ex.ac.uk/PErnest/) received a Ph.D. in philosophy of mathematics, but spent much of his career working in mathematics education. He is the editor of the *Philosophy of Mathematics Education Journal.* He has written an article, "The Impact of Beliefs on the Teaching of Mathematics" ([Ernest 1994]; originally written in 1989), suggesting that to make significant changes in mathematics education requires changing beliefs about the nature of mathematics as well as about how it is taught and learned. This then led to a book, *The Philosophy of Mathematics Education* [Ernest 1991]. More recently, he wrote a book setting forth his philosophical views of mathematics itself, *Social Constructivism as a Philosophy of Mathematics* [Ernest 1998]. As I make clear in my review of that book [Gold 1999], I do not view it as a viable version of social constructivism, but not all mathematicians agree with me, and I encourage readers to decide for themselves.

Annie and **John Selden** are the editors of the Research Sampler on MAA Online (http://www.maa.org/t_and_l/sampler/research_sampler.html), which brings selected research in mathematics education to the attention of collegiate mathematics educators. After long careers at various universities in the U.S. and abroad, they are now Adjunct Professors of Mathematics, New Mexico State University. In their own research, they have examined students' ability to solve novel calculus problems ([1989] and [2000]), students' grasp of the logical structure of informal mathematical statements, student difficulties with proofs [2003], and are currently investigating (college) teachers' beliefs about mathematics, teaching, and learning. See also their Research Questions page (http://www.maa.org/t_and_l/sampler/rs_questions.html) on MAA Online.

David Tall (http://www.warwick.ac.uk/staff/David.Tall/) also connects philosophical views of mathematics with his educational work in significant ways. See his "Existence Statements and Constructions in Mathematics and Some Consequences to Mathematics Teaching" [Tall/Vinner 1982], and a book he edited, *Advanced Mathematical Thinking* [1991]. More recently he has looked at the mathematical world as really three different realms [2004], the first coming from our perceptions of the world and thinking about them, the second the world of symbols we use in mathematics, the third the formal axiomatic world.

5.5 Mathematicians

I had a better success rate getting mathematicians who are interested in the philosophy of mathematics to contribute to this book. One who did not was **Saunders Mac Lane**; anyone interested in the philosophy of mathematics will find his *Mathematics: Form and Function* [1986] worth reading.

Chandler Davis has a very interesting view of mathematics, coming from a materialist perspective; see his "Materialist Mathematics" [1974] and "Criticisms of the Usual Rationale for Validity in Mathematics" [1990].

Lynn Steen has written a number of articles (and books) popularizing mathematics that include a philosophical bent. See particularly "The Science of Patterns" [1988].

Ian Stewart has also written a number of popular books about mathematics that have substantial philosophical implications. One of the best in that direction is *Nature's Numbers: The Unreal Reality of Mathematics* [1995].

6 A Brief Overview of This Book

This book is not designed for a straight read from beginning to end, although some readers might choose to do that. It is meant to be dipped into as the topic and writing style appeals to you. Each chapter is self-contained and most are liberally sprinkled with references for those wanting to delve more deeply into a particular topic. We have tried to organize it somewhat by topic, but within each topic the style and point of view of the chapters can be quite different. Thus, we've tried to provide you here with a guide to the chapters. Also each chapter is preceded by a short description of it and a brief biographical sketch of the author.

For mathematicians who have some curiosity about philosophical questions regarding mathematics, but who have not read any contemporary philosophy, a good place to start might be **Philip Davis**'s chapter. He asks a question that we have all wrestled with at some point, after we have done some mathematics and are thinking of writing it up: when is a problem solved? When can we say, OK, let's wrap that up now?

6.1 Views on Mathematical Objects

Barry Mazur's chapter should also be very accessible to mathematicians without much philosophical background. He asks a question that is close to a traditional philosophical question—how can one tell when one mathematical object is really the same as another. However, he looks in a very different direction than philosophers generally do as he traces some category theory from fundamentals to propose an answer this question, with some interesting comments along the way.

The others writing about mathematical objects are all philosophers.

Stewart Shapiro gives an overview of philosophical discussions concerning mathematical objects. This culminates with his view that mathematics is the science of structures (as suggested originally by Bourbaki), or, as it is sometimes called, of patterns.

For mathematicians interested in reading about current work in the philosophy of mathematics, **Charles Chihara**'s chapter is a relatively gentle introduction to the kind of discussions philosophers have. He discusses concerns about the existence of mathematical objects. Finally he turns to how one can develop a structural account of mathematics without being committed to the actual existence of structures in either the world or some ideal platonic realm.

Mark Balaguer gives an overview of the major variations philosophers have discussed over the last thirty years on whether there are mathematical objects; if there are, what their nature is; and how we can gain mathematical knowledge. As a summary of thirty years of philosophical discussion, this chapter is quite dense. However, it very effectively and systematically summarizes the discussion from a wealth of philosophical views.

Øystein Linnebo develops a new view of mathematical objects that allows them to exist in some sense while avoiding some of the traditional objections to a platonist account of mathematical objects.

6.2 Views on Proof

Proof and its relation to mathematical knowledge is an issue that has become an active concern again thanks to computer-assisted proofs and mathematical investigations involving computers.

Michael Detlefsen discusses both the role of empirical reasoning—primarily due to the use of computers—and of formalization in mathematical proofs.

Joseph Auslander discusses the various roles proof plays in mathematics, and how standards of proof vary over time.

Jon Borwein focuses less on proof than on the development of mathematics, and the roles computers may play in that development. Necessarily this includes the role they play in developing proofs.

6.3 What is Mathematics?

Robert Thomas's chapter suggests, as a definition of mathematics, a variation on "mathematics is the science of patterns." He takes mathematics as one extreme end in the spectrum of the sciences, and suggests (read the chapter for what he means) that "mathematics is the science of relations as such."

Guershon Harel approaches the problem of "what is mathematics?" from the viewpoint of a researcher in mathematics education. He proposes an answer that includes not only the theorems, but also the tactics and conceptualizations we use.

6.4 Social Constructivism

Reuben Hersh, one of the few mathematicians to attempt to describe social constructivism in some detail, discusses mathematics and its development (or, as he phrases it, mathematics as "a living organism") as the subject of scientific investigation. This lively chapter includes a beautiful attempt to describe the feeling when an idea for solving a problem suddenly flashes into one's mind.

Julian Cole just finished his Ph.D. thesis in philosophy, working on how one can make social constructivism coherent from a philosopher's standpoint. His chapter summarizes the main points of interest to mathematicians of his work.

6.4.1 The Boundaries Between Mathematics and the Other Sciences (Physical and Social), and the Applicability of Mathematics

Mark Steiner looks at a particular aspect of this question (primarily from the standpoint of a philosopher interested in the application of mathematics to physics) related to generalizations of addition.

Keith Devlin looks at the question of what we currently call mathematics versus what we currently relegate to applied mathematics. He describes how he believes this will change.

6.5 Philosophy of Probability

Alan Hájek discusses some of the fundamental philosophical issues about the nature of probabilities. It is a very accessible chapter.

Bibliography

[Benacerraf/Putnam 1964/1983] Benacerraf, Paul, and Hilary Putnam, eds., *Philosophy of Mathematics*, Cambridge: Cambridge University Press, 1st ed. 1964, 2nd ed. 1983.

[Benacerraf 1965] Benacerraf, Paul, "What Numbers Could Not Be," *The Philosophical Review*, 74 (1965), pp. 47–73.

[Benacerraf 1973] ——, "Mathematical Truth," *The Journal of Philosophy*, 70 (1973), pp. 661–680.

[Bishop 1967] Bishop, Errett, *Foundations of Constructive Analysis*, New York: McGraw-Hill, 1967.

[Brumbaugh 1954] Brumbaugh, Robert S., *Plato's Mathematical Imagination*, Bloomington IN: Indiana University Press, 1954.

[Burgess 1983] Burgess, John P., "Why I Am Not a Nominalist", *Notre Dame Journal of Formal Logic* **24** (1983), pp. 93–105.

[Burgess/Rosen 1997] Burgess, John P., and Gideon Rosen, *A Subject with No Object: Strategies for Nominalistic Reconstrual of Mathematics*, Oxford: Oxford University Press 1997.

[Burgess 2005] Burgess, John P., *Fixing Frege*, Princeton: Princeton University Press, 2005.

[Burton 2003] Burton, David M, *The History of Mathematics: An Introduction*, 5th edition, New York: McGraw Hill, 2003.

[Chihara 1973] Chihara, Charles, *Ontology and the Vicious-Circle Principle*, Ithaca: Cornell University Press, 1973.

[Corry 2004] Corry, Leo, *Modern Algebra and the Rise of Mathematical Structures*, Basel: Birkäuser, 2004.

[Davis 1974] Davis, Chandler, "Materialist Mathematics," pp. 37–66 in Cohen, Stachel and Wartofsky, eds., *For Dirk Struik, scientific, historical, and political essays in honor of Dirk J. Struik*, Boston: D. Reidel, 1974.

[Davis 1990] ——, "Criticisms of the Usual Rationale for Validity in Mathematics," pp. 343–356 in A.D. Irvine, ed., *Physicalism in Mathematics*, Dordrecht: Kluwer Academic Publishers 1990.

[Dubinsky 1986] Dubinsky, Ed, "Teaching mathematical induction I," *Journal of Mathematical Behavior* 5 (1986), pp. 305–317.

[Dubinsky 1989] ——, "Teaching mathematical induction II," *Journal of Mathematical Behavior* 8 (1989), pp. 285–304.

[Dubinsky et al. 1992] Breidenbach, D., E. Dubinsky, J. Hawks and D. Nichols, "Development of the process conception of function," *Educational Studies in Mathematics* 23 (1992), pp. 247–285.

[Dubinsky 1994] Dubinsky, Ed, "A theory and practice of learning college mathematics" pp. 221–243 in *Mathematical Thinking and Problem Solving*, A. Schoenfeld ed., Hillsdale: Erlbaum, 1994.

[Dubinsky et al. 1996] Asiala, M., A. Brown, E. Dubinsky, D. DeVries, D. Mathews and K. Thomas, "A framework for research and curriculum development in undergraduate mathematics education," pp. 1–32 in *Research in Collegiate Mathematics Education II, CBMS Issues in Mathematics Education*, American Mathematical Society, 1996.

[Dubinsky et al. 1997] Asiala, M., E. Dubinsky, D. Mathews, S. Morics, and A. Oktac, "Development of students' understanding of cosets, normality and quotient groups," *Journal of Mathematical Behavior* 16, 3 (1997), pp. 241–309.

[Ernest 1994] Ernest, Paul, "The Impact of Beliefs on the Teaching of Mathematics,", in Bloomfield, A. and T. Harries, eds., *Teaching and Learning Mathematics*, Derby: Association of Teachers of Mathematics, 1994.

[Ernest 1991] ——, *The Philosophy of Mathematics Education*, London: Falmer Press, 1991.

[Ernest 1998] ——, *Social Constructivism as a Philosophy of Mathematics*, State University of New York Press, Albany, NY, 1998.

[Feferman 1998] Feferman, Solomon, *In the Light of Logic*, New York: Oxford University Press, 1998.

[Fowler 1999] Fowler, David, *The Mathematics of Plato's Academy*, 2nd edition, New York: Oxford University Press, 1999.

[George/Velleman 2002] George, Alexander, and Daniel Velleman, *Philosophies of Mathematics*, Malden MA: Blackwell Publishers, Inc., 2002.

[Giaquinto 2002] Giaquinto, Marcus, *The Search for Certainty: A Philosophical Account of Foundations of Mathematics*, Oxford: Oxford University Press 2002.

[Gold 1999] Gold, Bonnie, review of *Social Constructivism as a Philosophy of Mathematics* and of *What is Mathematics, Really?*, *American Mathematical Monthly* 106, 4 (1999), pp. 373–380.

[Hersh 2005] Hersh, Reuben, *18 Unconventional Essays on the Nature of Mathematics*, New York: Springer-Verlag, 2005.

[Hesseling 2003] Hesseling, Dennis, *Gnomes in the Fog: The Reception of Brouwer's Intuitionism in the 1920s*, Birkhäuser Verlag, Basel, 2003.

[Hilbert 1919] Hilbert, David, "Natur u. mathematisches Erkennen", Lectures Fall semester, 1919.

[Körner 1968] Körner, Stephan, *The Philosophy of Mathematics: An Introductory Essay*, London: Hutchinson & Co., 1968.

[Lakatos 1976] Lakatos, Imre, *Proofs and Refutations*. Cambridge: Cambridge University Press, 1976.

[Mac Lane 1986] Mac Lane, Saunders, *Mathematics: Form and Function*, New York: Springer-Verlag 1986.

[Maddy 1990] Maddy, Penelope, *Realism in Mathematics*, Oxford: Oxford University Press 1990.

[Maddy 1997] ——, *Naturalism in Mathematics* Oxford: Oxford University Press 1997.

[Mancosu 2000] Mancosu, Paolo, "On Mathematical Explanation," pp. 103–119 in Emily Grosholz and Herbert Breger, eds., *The Growth of Mathematical Knowledge*, Dordrecht: Kluwer Academic Publishers, 2000.

[Manders 1989] Manders, Kenneth, "Domain extension and the philosophy of mathematics," *Journal of Philosophy* 86 (1989), pp. 553–62.

[Manders unpub.a] ——, "Why Apply Math?", unpublished manuscript.

[Manders unpub.b] ——, "Euclid or Descartes: Representation and Responsiveness", unpublished manuscript.

[Parsons 1979–80] Parsons, Charles, "Mathematical Intuition", *Proceedings of the Aristotelian Society* 80 (1979–80), pp. 145–168.

[Parsons 1983] ——, *Mathematics in Philosophy*, Ithaca: Cornell University Press, 1983.

[Putnam 1967] Putnam, Hilary, "Mathematics without Foundations", *Journal of Philosophy* 64 (1967), pp. 5–22.

[Putnam 1985] ——, *Mathematics, Matter and Method. Philosophical Papers*, vol. 1. Cambridge: Cambridge University Press, 1975. 2nd. ed., 1985.

[Resnik 1997] Resnik, Michael, *Mathematics as a Science of Patterns*, New York: Oxford University Press, 1997.

[Selden et al. 1989] Selden, A. and J., and A. Mason, "Can Average Calculus Students Solve Nonroutine Problems?" *Journal of Mathematical Behavior* 8 (1989), pp. 45–50.

[Selden et al. 2000] Selden, A., J. Selden, S. Hauk, and A. Mason, "Why Can't Calculus Students Access their Knowledge to Solve Non-Routine Problems?", *CMBS Issues in Mathematics Education*, 2000.

[Selden 2003] Selden, A. and J., "Validations of Proofs Considered as Texts: Can Undergraduates Tell Whether an Argument Proves a Theorem?" *Journal for Research in Mathematics Education* 34(1) (2003), pp. 4–36.

[Steen 1988] Steen, Lynn, "The Science of Patterns," *Science* 240 (29 April 1988) pp. 611–616.

[Stein 1988] Stein, Howard, "Logos, Logic, and Logistiké: Some Philosophical Remarks on 19th Century transformation of Mathematics," pp. 238–259 in William Aspray and Philip Kitcher, eds., *History and Philosophy of Modern Math*, Minneapolis: University of Minnesota Press, 1988.

[Stein 1989] ——, "Yes, but . . . : Some Skeptical Reflections on Realism and Anti-realism," *Dialectica* 43 (1989), pp. 47–65.

[Stein 1990] ——, "Eudoxos and Dedekind: On the Ancient Greek Theory of Ratios and its Relation to Modern Mathematics," *Synthese* 84 (1990), pp. 163– 211.

[Stewart 1995] Stewart, Ian, *Nature's Numbers: The Unreal Reality of Mathematics*, Perseus Publishing, 1995.

[Tait 2005] Tait, William, *The Provenance of Pure Reason: Essays in the Philosophy of Mathematics and Its History*, New York: Oxford University Press, 2005.

[Tall/Vinner 1982] Tall, David, and Shlomo Vinner, "Existence Statements and Constructions in Mathematics and Some Consequences to Mathematics Teaching," *American Mathematical Monthly* 89, 10 (1982), pp. 752– 756.

[Tall 1991] Tall, David, ed., *Advanced Mathematical Thinking*, Kluwer: Holland 1991.

[Tall 2004] Tall, David, "Introducing Three Worlds of Mathematics," *For the Learning of Mathematics* (2004).

[Tieszen 2005] Tieszen, Richard, *Phenomenology, Logic, and the Philosophy of Mathematics*, Cambridge: Cambridge University Press, 2005.

[von Neumann 1966] von Neumann, John, *Theory of Self-Reproducing Automata*, A.W. Burks, ed., Urbana: University of Illinois Press, 1966.

I

Proof and How it is Changing

Proof has been an essential part of mathematics since the time of the ancient Greeks. Its centrality has engendered much controversy. What is the role of proof in mathematics? What makes for an adequate proof?

The recent use of computers in developing mathematical conjectures, and in checking cases when there are too many for humans to check in a reasonable amount of time, has led to questions about the role and importance of proof in mathematics, as well as what qualifies as a proof. The chapters in this section give three different views of these and other issues regarding relationships among proof, mathematics, and computers.

1

Proof: Its Nature and Significance

Michael Detlefsen
Professor of Philosophy
University of Notre Dame

From the Editors

In our first chapter, Michael Detlefsen carefully examines the historical tension between inductive and deductive methods in mathematics, and relates it to the current discussion of the roles of each in the development of mathematics. He then turns to the question of whether, in fact, formalization of proofs actually increases either understanding or reliability of proofs. He also summarizes recent work on diagrammatic reasoning in mathematics, and the possible roles of visual experience in proofs.

We have chosen this as the first chapter in the book because we believe it is a fine, careful examination of these questions that virtually every reader of this volume will benefit from reading. For those of us who teach mathematics, an awareness of the fluctuations in the role of proof, and what is considered a proof, can be of use in the classroom. Such awareness can give us both a context in which to set our students' attempts at proof and a historical background we can impart to our students. An awareness of the importance of inductive methods in the development of mathematics is also worth transmitting to our students. In particular, making students aware of the current discussion in the mathematical community about the role of computers in mathematics can help them realize that mathematics is still a growing subject, even if most of the mathematics they study at the undergraduate level is centuries old.

Michael Detlefsen is a Professor of Philosophy at the University of Notre Dame (philosophy.nd.edu/people/all/profiles/detlefsen-michael/). His interests include logic, the philosophy of mathematics, and more specifically the role of proof in mathematics. He has written one book, Hilbert's Program *(1986), and edited two others,* Proof, Logic, Formalization *(1991),* Proof and Knowledge in Mathematics *(1991). Among his articles that are likely to be of particular interest to readers of this volume are "The Four-Color Theorem and Mathematical Proof" in* The Journal of Philosophy *(1980), "Poincare vs. Russell on the Role of Logic in Mathematics,"* Philosophia Mathematica *(1993), "Mind in the Shadows: Essay Review of Roger Penrose's The Emperor's New Mind (OUP, 1989), Shadows of the Mind (OUP, 1994) and The Large, the Small and the Human Mind (CUP, 1997),"* Studies in the History and Philosophy of Modern Physics *(1998), "What does Gödel's Incompleteness Theorem Say?"* Philosophia Mathematica *(2001), and "Formalism," in* The Oxford Handbook of the Philosophy of Mathematics and Logic *(2005). He's currently working on "The Role of the Imaginary in Mathematics" and has a forthcoming article, "Purity as an Ideal of Proof," to appear in* The Philosophy of Mathematical Practice.

1 Introduction

Recent philosophical work on the topic of mathematical proof has focused on epistemological concerns. Prominent among these are the questions whether

(i) there is a special type of knowledge that proof and proof alone supports, or for which it provides special support,

whether

(ii) the knowledge supported by proof warrants a regimentation of mathematical practice that makes proof the sole legitimate or at least the preferred form of justification in mathematics

and, relatedly, whether

(iii) there is a place for broadly empirical reasoning in the development of mathematical knowledge.

These concerns are not new, of course, but have been of perennial interest to mathematicians and philosophers. Traditionally, responses to (i) have generally been affirmative. Views on (ii) and (iii) have been more mixed, with some arguing that empirical reasoning has little if any place in the development of mathematical knowledge and others (in roughly equal numbers) maintaining that it plays a vital role.

For most of the past three centuries, philosophical work on mathematics has mainly admitted the usefulness of empirical methods while also insisting that they do not provide the same quality of knowledge as classical demonstration or proof.

Recent work has, for the most part, sustained this compromise. It has, in particular, supported the use of various types of empirical reasoning to help solve mathematical problems. Yet while earlier thinkers generally based their support of such methods on considerations of usefulness, convenience or perhaps practical necessity, recent writers have sometimes strengthened this to something verging on psychological or perhaps even physical necessity.

This is nowhere more evident than in the recent controversy concerning the status of the computer-assisted resolution of the four-color problem in 1977. Here some have argued that the proof of the four-color theorem (or any other extremely long and/or complex argument) represents a fundamental departure from traditional standards and methods of justification in mathematics. Others have argued the contrary. My sympathies are primarily with the latter.

An examination of the historical record reveals, I think, that there have long been arguments recommending the use of inductive arguments in mathematics on the grounds of their usefulness and practical necessity. I survey some of this history in section 2 and relate it to more recent work in section 3.

In section 4, I consider a different challenge to the currently prevailing view of proof—one which focuses on rigor and the conditions necessary for its attainment. Of particular interest there is a recent series of papers by the artificial intelligence researcher J. A. Robinson, who argues that proof has two essential aims. One is to convince, the other to explain. He observes that though formalization may assist the convictive aim of proof in certain ways, it can also obstruct its explanatory aim. This being so, it may compromise the greater ends of proof. It may even interfere with rigor since, as he maintains, explanatory coherence is sometimes the best protection we have against serious gaps in our reasoning.

Robinson's overall goal is to explain a notable phenomenon—namely, the apparent gap that exists between the standards of proof that seem to guide mathematical practice, and the more austere standards of formalization that prevail in mathematicians' descriptions of their ideals. If formalized proof is indeed the ideal of mathematical justification, why should ordinary practice remain so far from it? The common response would be that the two are not so remote from each other, that formalization of ordinary proof is largely a routine affair. Robinson challenges this view both with arguments and examples. More positively, he proposes a conception of proof which emphasizes its affinity with performance rather than pure text. The upshot is a view that promises to be at once subtler and empirically more realistic than the currently prevailing views of proof and rigor.

In section 5, I turn to the large, wide-ranging recent literature on diagrammatic reasoning and its place in mathematics. This intersects with the topics treated earlier in that it stresses the role of visual experience in proof and considers how mathematical thinking might make use of such experience while still remaining properly rigorous. Of particular interest in this connection is the radical view of Jon Barwise and his co-author John Etchemendy (and their students), who claim that diagrammatic reasoning can play not only a heuristic but a genuinely justificative role in proof. At the heart of their view is the belief that visual and linguistic representations of the same information can and often do have significantly different properties. In particular, they commonly differ with respect to certain types of efficiency—the diagrammatic variants being generally more efficient in these ways. Describing and accounting for such differences has been a major preoccupation of both their work and other recent work on diagrammatic reasoning.

So too has been the question of how visual and/or diagrammatic reasoning fits with the explanatory goals of proof and the quest for rigor. Of particular interest in this connection is a body of work by the philosopher Marcus Giaquinto, who, more than anyone else, has taken pains to clarify both the senses in and the extent to which diagrammatic reasoning figures in justificatively significant ways in mathematical reasoning. He refines the description of the types of justificative contributions diagrammatic reasoning can make and carefully investigates its justificative limits, particularly in analysis. Finally, he considers the difficult question of how diagrammatic reasoning fits with the explanatory aims of proof.

In section 6, I summarize and conclude. I find that much recent work continues the dominant view of the last three centuries in its view of the place of empirical reasoning in mathematics. The chief novelties concern refinements in our understanding of the nature and role(s) of diagrammatic reasoning and of the proper place of formalization in proof.

2 Empirical Reasoning in Mathematics: Historical Background

The use of empirical evidence and broadly inductive reasoning in mathematics are by no means new phenomena in the history of mathematics. They have, in fact, been a major source of concern for more than two millenia. Archimedes gave an important early defense of the usefulness of empirical methods (particularly, mechanical methods) in solving geometrical problems. He did not see them as altogether supplanting classical methods, but he did see them as useful means of discovery (both of truths and of proper demonstrations).

> ... I have thought fit to ... explain ... a method, with which ... you[1] will be able to make a beginning (*aphormē*) in the investigation (*thēorein*) by mechanics ... in mathematics.... [I]nvestigation by this method does not amount to actual proof (*apodeixeōs*); but it is ... easier to provide the proof when some knowledge of the things sought has been acquired by this method rather than to seek it with no prior knowledge.
>
> [Archimedes], pp. 221, 223[2]

Archimedes thus acknowledged a role in mathematics for reasoning other than proof. Though proof might yield knowledge in its highest form(s), other forms of reasoning might nonetheless yield lesser knowledge, and also aid in the development of genuine proofs.

It was thus common for non-demonstrative methods of reasoning to be classified as methods of invention or discovery, as distinct from methods of justification proper. The terminology is potentially misleading, though, in that it suggests that non-demonstrative reasoning was not viewed as justificative. This is not true. Discovery of a proposition was discovery of its truth. Discovermental methods were thus generally taken to have justificative value, but not so great as that typical of demonstration. This reflected the Aristotelian "causal" ideal of knowledge.

> We ... possess unqualified scientific knowledge of a thing ... when we know the cause on which the fact depends as the cause of the fact and ... that the fact could not be other than it is. [Aristotle 1908], 71b8–b11

The broad division of reasoning into justificative and discovermental varieties is thus of ancient origin. Philosophers and mathematicians generally marked it and so too did other disciplines in which reasoning featured prominently. As Cicero remarked: 'every careful method of arguing has different divisions—one of discovering, one of deciding' ([Cicero 1894], vol. 4, *Topics*, pp. 459–460)).[3] Methods of the former type were termed *arts of discovery* (*artis inveniendi*), methods of the latter type *arts of justification* (*artis iudicandi*).

16th and 17th century algebraists embraced the distinction. Viète, Descartes and Wallis all stressed the different purposes served by discovery-oriented and demonstrative reasoning and the different standards to which they are rightly held. They also saw in it a reflection of the ancient

---

[1] The person addressed was Eratosthenes.

[2] The reader should bear in mind that the text for Archimedes' *Method* was only rediscovered at the turn of the 20th century and cannot generally be assumed to have been available to earlier thinkers.

[3] Cicero attributed the distinction to Aristotle, who urged a similar division in the *Topics*. It was also marked in Roman law, which distinguished evidence appropriate to the detention, questioning and/or charging of a suspect (*investigatio*), from evidence appropriate to her conviction (*demonstratio*).

distinction between analysis and synthesis. The classical statement of this distinction was given by Pappus of Alexandria in his *Treasury of Analysis*.

> . . . in analysis we suppose that which is sought to be already done, and inquire what it is from which this comes about . . . until, by retracing our steps, we light upon something already known or ranking as a first principle . . .
>
> . . . in synthesis . . . we suppose to be already done that which was last reached in analysis, and arranging in their natural order as consequents what were formerly antecedents and linking them one with another, we finally arrive at the construction of what was sought . . . [Pappus 2000], pp. 597, 599[4]

They believed, moreover, that symbolic algebra (what they commonly referred to as the *analytic art*) was analysis *par excellence*. It was, in the first instance, a method of discovery widely believed to be efficient and reliable. Wallis described it as 'plain, obvious and easy' ([Wallis 1685], p. 298) and as yielding results that were readily verifiable by classical means (cf. [Wallis 1685], p. 305). In a similar spirit, Leibniz praised it as "a great aid in shortening thought and also in discovery" ([Leibniz 1707], p. 436) and claimed that it could not "lead us into error" (*loc. cit*).

Others (e.g. MacLaurin, cf. [MacLaurin 1742], pp. 47, 49) conceded its usefulness but also emphasized that it was not the justificative equal of classical (synthetic) method.

> In general, it must be owned, that if the late discoveries [of Wallis', in his *Arithmetica Infinitorum* (1656)] were deduced at length, in the very same method in which the ancients demonstrated their theorems, the life of man could hardly be sufficient for considering them all . . . [MacLaurin 1742], p. 49, brackets mine[5]

Still, though

> [m]athematicians [may] indeed abridge their computations by the supposition of infinites, . . . [they] cannot be too scrupulous in admitting of infinites, of which our ideas are so imperfect. [MacLaurin 1742], pp. 46–47, brackets mine

Still others less qualifiedly opposed the use of algebraic methods. These included Hobbes ([Hobbes 1839–45], vol. 1, pp. 311–312)[6] and, at times, Newton.

> Equations are Expressions of Arithmetical Computation, and properly have no Place in Geometry . . . Multiplications, Divisions, and such sort of Computations, are newly received into Geometry, and that unwarily, and contrary to the first Design of this Science. [Newton 1720], p. 229

---

[4] Aristotle made a similar distinction earlier [Aristotle 2000] (cf. III.3, 1112b 15–27). There is also a statement of uncertain origin in the manuscript sources for Book XIII of Euclid's *Elements*. See Heath's historical note on Book XIII ([Euclid 1956], vol. 3, pp. 438–439 and his commentary on Propositions I–V ([Euclid 1956], vol. 3, pp. 441–442) for more on this.

[5] For similar statements during the same period, see that by Christian Wolff ([Wolff 1739], preface, v–vi). A representative statement a half century earlier was that by Johann Christoph Sturm ([Sturm 1700], preface, articles XIV, XX), a half century later that by Charles Hutton ([Hutton 1795], vol. 1, p. 107).

[6] Hobbes and Wallis carried on a well-known dispute concerning the legitimacy of algebraic methods. An interesting account of this dispute is given by Jesseph in [Jesseph 1999].

To counter such charges, algebraists of the 16th and 17th centuries argued that classical geometry would not have been the success it had been had not ancient geometers made regular use of algebraic methods in arriving at their discoveries—a use they then tried to conceal (cf. Descartes [Descartes 1620–28], Rule IV, [Wallis 1685], ch. II, 3, p. 290 and Viète [Viète 1591], p. 27). Wallis thus wrote of Apollonius that

> ... we may well give him the name of *Magnus Geometra*, and look upon him as a man of a prodigious reach of Phansy, if we can think it possible that he could discover all those Propositions, and perplex demonstrations, in the same order they are there delivered, without some such Art of Invention, as what we now call Algebra.
>
> [Wallis 1685], p. 290

This is the traditional view of analytic or algebraic method and its place in classical geometry. But why classify it as empirical in character? Generally speaking, the reason is that it often relied on inductive forms of reasoning. Archimedes and various other ancient and medieval mathematicians appealed to analogies between mechanics and geometry, and algebraists of the early modern era (i.e. late 15th–17th centuries) often used a form of inductive reasoning they associated with the *Principle of Exhaustion*—the idea that if two quantities can be made to differ from each other by less than any assignable amount, they can then be treated as equal.[7]

They often reasoned by analogy as well, extending laws proven for finite magnitudes and collections to infinite generalizations of them. One example of this is Wallis' extension of the law for sums of (finite) arithmetic progressions to sums of infinite arithmetic sequences ([Wallis 1656], p. 155; [Wallis 1685], pp. 285–287, 297, 305–306). Overall the method was inductive and Wallis described it as such.

> The simplest method of investigation, in ... various problems ... is to exhibit the thing to a certain extent, and to observe the ratios produced and to compare them to each other; so that at length a general proposition may become known by induction.
>
> [Wallis 1656], p. 13

Following this procedure, Wallis "retrieved" the classical law for the area of the triangle (viz. $A = \frac{bh}{2}$)[8] by applying Cavalieri's Method of Indivisibles. He resolved the triangle into a progression of uniformly thin rectangles, reasoned inductively that, compositely, they would come ever closer to matching the interior of the triangle as the individual rectangles became ever thinner, and reasoned analogically to their sum by extending the formula for arithmetic progressions to the "infinite" case.[9] He thus arrived at $\frac{0\times\frac{h}{\infty}+b\times\frac{h}{\infty}}{2} \times \infty$, thence $\frac{\frac{bh}{\infty}\times\infty}{2}$, thence $\frac{bh}{2}$, the classical law for the area of a triangle.[10]

Wallis repeated the same general form of reasoning—finding the appropriate type of progression and analogically inferring its sum—to obtain solutions to a variety of other quadrature

---

[7] The Principle of Exhaustion was well-known and widely used in antiquity (cf. Def. IV, Bk. V of *The Elements*, which is used to prove another variant in Prop. I, Bk. X). Democritus is generally thought to have been the first to formulate it, though there is evidence that Hippocrates formulated and used it too. Archimedes attributed the first "proof" of it to Eudoxus.

[8] Cf. Wallis [Wallis 1656], Proposition 3; [Wallis 1685], pp. 285–287.

[9] Cf. Wallis [Wallis 1656], pp. 13–15; [Wallis 1685], pp. 280–290.

[10] Cf. Wallis [Wallis 1656], pp. 14–15, [Wallis 1685], pp. 285–287.

and cubature problems as well.[11] There was immediate, sharp criticism from both philosophers and mathematicians, most notably Hobbes, Huygens and Fermat. All three criticized Wallis for his use of inductive reasoning, characterizing it variously as unclear, uncertain, unnecessary and insufficient.[12] Wallis' chief response, stated in a reply to Fermat, was to reaffirm the ancient *two-methods* (discovery vs. demonstration) distinction. His aim, he said, was not primarily one of "Demonstrating things already known" ([Wallis 1685], p. 305), but "to shew a way of... finding out... things yet unknown" (*ibid.*).

Thus Wallis, and algebraists of 16th and 17th centuries generally, embraced the ancient distinction between an *ars inveniendi* and an *ars iudicandi*. In the next section we'll consider a recent proposal by Arthur Jaffe and Frank Quinn to divide mathematical labor in a roughly similar way between discovermental and more rigorously demonstrational components. We'll consider as well a widely-discussed argument (concerning the computer-assisted proof of the four-color theorem) that empirical reasoning may sometimes be the only humanly feasible means of justification.

3 Empirical Reasoning in Mathematics: Recent Proposals

3.1 Empirical Reasoning and Epistemic Productivity

Arthur Jaffe and Frank Quinn recently offered a new incarnation of the division of methods theme. They acknowledge the benefits of rigor as a constraint on mathematical reasoning and so affirm the virtues of strict proof. It has 'brought to mathematics a clarity and reliability unmatched by any other science' ([Jaffe/Quinn 1993], p. 1). This notwithstanding, it has also at times made progress in mathematics 'slow and difficult' (*loc. cit.*).

Too strong an emphasis on proof may thus be more of an impediment than an aid to the development of new mathematical knowledge. To become more efficient, they suggest, mathematics should follow the lead of physics and permit freer use of intuitive methods of thinking. And this despite the fact that by means of such more liberal reasoning, mathematicians may occasionally go beyond the bounds of what can be strictly established (*op. cit.*, p. 2).

Freer, more efficient 'theoretical' methods[13] should be used to generate initial hypotheses and to outline justifications. These hypotheses and justifications should then be converted into rigorous reasoning by mathematicians particularly skilled in such work.

In Jaffe and Quinn's view, the role of rigorous proof in mathematics is 'functionally analogous to the role of experiment in the natural sciences' (*loc. cit.*). They thus foresee two types of mathematical research—a more intuitive and speculative 'theoretical' type aimed at efficient discovery, and a more rigorous, conventional type aimed essentially at confirmation. The latter is intended (i) to 'ensure the reliability of mathematical claims' (*loc. cit.*), and (ii) to yield, at least occasionally, 'new insights and unexpected new data' (*loc. cit.*).

---

[11] Cf. Wallis [Wallis 1656], *passim*; [Wallis 1685], pp. 285–290, 290–298.

[12] There were admirers too, of course. In addition to those mentioned above, these included Newton (at times) and, nearly two centuries later, Charles Babbage. Such influential admirers notwithstanding, Wallis' inductive and analogical methods did not change the norms of mathematical practice. Succeeding generations of mathematicians for the most part viewed them as falling short of ideal norms of rigor, certainty and precision.

[13] So called because they resemble thinking in *theoretical* physics.

Dividing mathematics in this way, they suggest, may bring about the same rapid advancement in it as it did in physics. This, at any rate, is their hope. Their proposal goes farther, however, in proposing that the division of methods be incorporated into the institutions of professional mathematics, specifically, into its methods of training and its system of rewards.

3.1.1 The Division of Mathematical Labor

Fundamentally, Jaffe and Quinn's proposal is one of divided labor and, as such, it embodies the same strategic ideas that schemes of divided labor generally embody. Specifically, it proposes to increase productivity through increased specialization.[14]

Adam Smith's classical statement of the benefits of divided labor maintained that it 'occasions, in every art, a proportionable increase of the productive powers of labor' ([Smith 1776], Bk. I, ch. 1, para. 4). By dividing production into small tasks and 'dedicating' each individual worker to the repeated execution of a single task (or a small number of such), productivity is increased. This is so because the tasks are 'smaller' and, so, more fully within the range of the worker's competence, and because the familiarity that comes from repetition increases the worker's proficiency in performing them. The result is a better product more efficiently produced. Or so the thinking goes.

At bottom, this is what Jaffe and Quinn place their faith in. Mathematics will be divided into specialists in speculation or conjecture (practitioners of 'theoretical' methods in mathematics) and specialists in confirmation (those who convert 'theoretical' reasonings into proofs). With this increased specialization will come increased proficiency, and with increased proficiency, increased productivity. Mathematical knowledge will both improve in quality and grow faster.

There are grounds for caution, however. One is the general lack of evidence for the claims and assumptions that Jaffe and Quinn make. Perhaps the most basic of these is the assumption that the recent growth of knowledge in physics is greater than that in mathematics. Even granting this assumption, though, questions remain. For example, do we know that it's the division of physics researchers into theoretical and experimental that's responsible for its superior rate of growth? And, supposing that it is, is what makes that division effective its separation of the speculative (roughly discovermental) and confirmatory (roughly justificative) tasks? Or might it instead be the increase in financial support for physics (and applied mathematics) research fueled by the race to develop atomic weapons and energy, or the race to put humans on the moon? Is the task of developing a proof for a conjecture in mathematics relevantly similar to that of confirming a physical conjecture? Is it generally as *easy* to devise a proof for a true mathematical conjecture as it is to design and conduct a confirmatory experiment (or body thereof) for a true physical conjecture? Such questions are not easily answered.

In addition to these uncertainties, there is another that may be of even greater significance. It has to do with the costs of dividing labor. Even Smith, the champion of divided labor, acknowledged these costs and that they are considerable.

> In the progress of the division of labour, the employment of the far greater part of those who live by labour ... comes to be confined to a few very simple operations,

[14] It may also be that Jaffe and Quinn believe something like what Archimedes expressed when he claimed that it's 'easier to provide the proof when some knowledge of the things sought has been acquired' ([Archimedes], p. 223). The informal sketches of justifications produced by 'theoretical' mathematics may give the rigorist something to directly build on.

> frequently to one or two.... The man whose whole life is spent in performing a few simple operations, of which the effects are perhaps always the same, or very nearly the same, has no occasion to exert his understanding or to exercise his invention in finding out expedients for removing difficulties which never occur. He naturally loses, therefore, the habit of such exertion, and generally becomes as stupid and ignorant as it is possible for a human creature to become. [Smith 1776], Bk. V, ch. 1, para. 178

A more disheartening view of the effects of divided labor would be hard to imagine. Consistent division of labor, on this view, impedes the worker's ability to find fulfillment in her work. In short, it alienates her from her work.

In addition to concerns regarding the accuracy of their comparison of mathematical and physical research, then, their proposal also raises larger moral and social concerns that have not been adequately addressed.

3.2 Empirical Reasoning vs. Proof

We saw in section 2 how broadly inductive reasoning and reasoning from analogy have been used as methods of discovery in mathematics since ancient times. In all cases, however, discovery and justification (at least ultimate or ideal justification) were conceived as distinct tasks. Discovermental arguments were therefore to be only temporary substitutes for proper demonstrations.

In recent times a different role for empirical reasoning in mathematics has been suggested. It is no longer seen as a mere propædeutic to proof, but an alternative to it—in some cases, a necessary alternative.

Views of this type have been inspired by the appearance of extremely long and/or complex proofs. A well-known example is the widely discussed computer-assisted solution of the four-color problem developed by Kenneth Appel and Wolfgang Haken in 1977.

This proof is so large as to seemingly prohibit the type of step-by-step surveyal commonly assumed for mathematical proof. It has thus given rise to a variety of philosophical questions concerning whether proof is indeed the appropriate standard to adopt for mathematical justification.

Thomas Tymoczko wrote a widely read discussion of these issues in his 1980 paper 'The Four-Color problem and its mathematical significance'. He argued that the solution of the four-color problem offered by Appel and Haken (hereinafter, the AH argument or proof) forced a reconsideration of traditional conceptions of mathematical proof and knowledge. His main claim was that

> ... if we accept the 4CT as a theorem, we are committed to changing the sense of 'theorem', or, more to the point, ... the sense of the underlying concept of "proof".
>
> ... use of computers in mathematics, as in the 4CT, introduces empirical experiments into mathematics. Whether or not we choose to regard the 4CT as proved, we must admit that the current proof is no traditional proof, no a priori deduction of a statement from premises. It is a traditional proof with a ... gap, which is filled by the results of a well-thought-out experiment. This makes the 4CT the first mathematical proposition to be known a posteriori and raises again for philosophy the problem of distinguishing mathematics from the natural sciences. [Tymoczko 1979], p. 58

Tymoczko's argument had three main components. The first was an analysis of the traditional conception of proof which identified three key characteristics—convincingness, surveyability and formalizability. The second was an argument to the effect that, of the three ingredients just mentioned, surveyability was the most basic. The third was the claim that the AH proof is not surveyable.

I'll now consider the argument more carefully, focusing, as Tymozcko did, on the two latter components—the claims that surveyability is central to the standard conception of proof and that the AH proof is not surveyable.

In calling a proof convincing, Tymoczko meant that it had the capacity to move a rational prover[15] to belief in its conclusion. This in turn required that the proof be surveyable—that is, that it be capable of being 'looked over, reviewed, verified' (*op. cit.*, p. 59) by human provers, specifically, by members of the human mathematical community (*op. cit.*, p. 60; [Tymoczko 1980], p. 132).

Such a conception of proof was, in Tymoczko's view, seriously at odds with the computer-assisted proof of the 4CT. This proof established the existence of a formalized proof of the 4CT, a formalized proof so large, however, as to debar human survey (cf. [Tymoczko 1979], p. 58). The AH proof of the 4CT thus substituted, at certain point(s), the *results* of unsurveyably long computer runs for the formal computations to which they correspond. So, at any rate, Tymoczko claimed.

To evaluate this reasoning, we need to keep two distinct items separate. One is the actual argument given—that is, written down—by Appel and Haken. This argument is surveyable and was indeed surveyed by a number of mathematicians before being published. I'll call this the *compressed argument*, since it replaces the details of certain computations with (descriptions of) their results.

The unsurveyable argument, on the other hand, is the argument that would result from setting out the suppressed details of the compressed argument in full. Call this the *decompressed argument*.

Judged by Tymoczko's standards, neither the compressed nor the decompressed argument is a proof in the traditional sense. The decompressed argument is not a proof because it is not surveyable. The compressed argument is not a proof because it lacks the explicitness—the full disclosure of premises and inferences—traditionally required of proof. It is

> ... like a mathematical proof where a key lemma is justified by an appeal to the results of certain computer runs or, as we might say "by computer." This appeal to computer, whether we count it as strictly a part of a proof or as a part of some explicitly non-proof-theoretic component of mathematical knowledge, is ultimately a report on a successful experiment. It helps establish the 4CT (actually, the existence of a formal proof of the 4CT) on grounds that are in part empirical. [Tymoczko 1979], p. 63, brackets added

In Tymoczko's view, then, to accept the compressed argument (i.e. the AH proof) as adequate mathematical justification for the 4CT requires changing the traditional conceptions of proof and

[15] By a 'prover', I mean not only one who discovers a proof, but also one who grasps it.

mathematical knowledge (cf. [Tymoczko 1979], p. 58). In particular, it requires relinquishing or modifying each of the following (cf. *op. cit*, p. 63):

(i) all mathematical theorems are known *a priori*
(ii) mathematics has no empirical content
(iii) mathematics relies exclusively on proof and makes no use of experiment, and
(iv) mathematical theorems are certain to a degree that no theorem of natural science can match.

Tymoczko thus took the acceptance of the AH argument for the 4CT to be both a novel and a philosophically significant development.

3.3 The Philosophical Significance of the AH Argument

Items (i)–(iii) center on the question of the compatibility of the commonly supposed *a priori* status of mathematical judgments with the use of broadly empirical considerations in mathematical reasoning. To get a better idea of the plausibility of Tymoczko's claim that (i)–(iii) must be relinquished or modified, we must first get clearer on what he means (or might or should mean) by 'mathematical theorem', 'known *a priori*', 'empirical content' and 'empirical justification'.

Tymoczko assumes that the traditional view of theorems is essentially this:

(i-aux): A proposition is rightly classified as a theorem if and only if there is a known proof of it.

From (i-aux) and the supplementary claim that

(i-supp): If there is a known proof for a proposition, then it (the proposition) is known *a priori*,

(i) follows.

In truth, though, the claim that (i-aux) represents the traditional view is doubtful, at least if it is taken to imply that a known proof is a proof that has been surveyed in all its details. There is nothing in the traditional conception of proof to prohibit joint enterprises where a resolutive task is broken up into parts, and the various parts given over to different persons or groups in such a way that, in the end, no one participant will have surveyed the entire joint proof. The product of such an undertaking could still count as a proof so long as it was known that each part was correctly executed and that, taken together, the several parts solve the original problem.[16]

If this is right, a more reasonable condition than (i-aux) would be

(i-aux'): A proposition is rightly classified as a theorem if and only if it is known to have a proof.

(i-aux'), however, implies (i-aux) only on a constructive understanding of existence, and such an understanding is not part of the traditional conception of proof.

[16] A recent well-known example of such a joint undertaking is the classification of the finite simple groups.

Even if it were, though, problems would remain. For it's only the stricter forms of constructivism that require actual exhibition (or survey) of an object as adequate justification of its existence. More liberal varieties allow existence to be established by the provision of suitably clear descriptions of methods the (perhaps idealized) execution of which can be seen to guarantee an exhibition of a thing of the type claimed to exist.

(i-aux) thus has little to recommend it as a traditional condition of theoremhood. Related remarks apply to (i-supp) and condition (i). (i-supp) must give way to

(i-supp'): A proposition known to have a proof is known to have an *a priori* justification[17],

and (i) to

(i') A proposition rightly classified as a theorem is known to have an *a priori* justification.[18]

The question then becomes whether the compressed argument for the 4CT comports with (i'), and it seems that it does. It (the compressed argument) can reasonably be taken to show that there is a proof of the 4CT, and this together with (i-supp') implies that the 4CT has an *a priori* justification. This being so, the compressed argument would not seem to demand a revision of what is in truth the traditional view—namely, that whether or not they are humanly graspable, there nonetheless exist *a priori* justifications for mathematical theorems.

Let's now consider (iii). As stated, it's too vague to assess. Clarified in the way Tymoczko's suggests, however, it runs counter to the traditional conception of proof. This conflict is due to a strong property of self-sufficiency that Tymoczko attributes to the traditional conception of proof. A proof is, he says,

> ... an exhibition, a derivation of the conclusion, and it needs nothing outside of itself to be convincing. [Tymoczko 1979], p. 59

Is such self-sufficiency characteristic of the traditional conception of proof?

There are reasons to think that it isn't. Conviction often, perhaps typically, requires not only proofs *per se* but *reflections on* them. These may be as simple as the application of certain checking procedures or as complex as reflections on the meaning and/or plausibility of ideas and principles used in a proof. Whatever their particular character, reflections on proofs and their components are often vital to acceptance of their conclusions. At the same time, though, they do not strictly belong to the proofs themselves, at least not as proofs are ordinarily thought of these days.[19]

Paul Teller ([Teller 1980]) made a similar point in arguing that Tymoczko was wrong to regard surveyability as a necessary condition of proof. It's not a property of proofs *per se*, he said, but a property some proofs have and others lack. It signifies not the extent to which an argument actually *is* a proof, but the extent to which it can be verified as such. It's thus a matter of degrees. Some proofs are so simple and easy that little knowledge and training is required

[17] Briefly, a proposition p will be said to *have* an *a priori* justification if there is a warrant for it that does not depend on taking the contents of any experience as evidence.

[18] Later we'll see reason to think that this might even be strengthened to something like 'A proposition known to have a proof is known *a priori*'.

[19] This is in part due to the fact that modern understandings of the notion of *axiom* do not typically retain the classical requirement of self-evidence.

for their verification. Others 'are so complicated that only a few mathematicians' ([Teller 1980], p. 798) can verify their correctness. Still others are 'out of the reach of the best [verifiers]' (*loc. cit.*). Despite this variation in their verifiability, however, all may be genuine proofs. The AH proof of the 4CT thus represents at most an extension of our means of surveying proofs, not a change in our concept of proof.

All in all, it seems wrong to say that, on the traditional conception, proof is a unit of reasoning that needs nothing outside itself to be convincing. (iii) is therefore not a basic tenet of the traditional conception of proof. Conviction in mathematics often involves not only proofs but judgments *about* them—judgments which do not themselves belong to the proofs in question. Neither has conviction traditionally been seen as requiring proof, as the discussion of section 2 makes clear. In sum, proof has not traditionally been regarded as either necessary or sufficient for conviction.

Tymoczko and his claims aside, the AH proof of the 4CT raises other interesting questions regarding proof and mathematical knowledge. One of these concerns what if any difference there might be between actually having an *a priori* warrant for a proposition and simply having evidence that one exists.

Suppose that I become convinced, on evidence I know to be reliable, that a certain statement κ is provable from certain *a priori* warranted statements $\pi_1, \ldots, \pi_n$ whose warrants I know to be reliable. Suppose, in addition, that I have good reason to believe that the shortest proof of κ from these statements is not humanly surveyable.

Under such circumstances, it would not be humanly possible to have a proof of κ from the statements mentioned, and I might even know or believe this to be so. This notwithstanding, I might still know that *there is* such a proof. More exactly, under the circumstances described, my knowledge that the statements in $\pi_1, \ldots, \pi_n$ are *a priori* warranted, and my knowledge that κ is provable from $\pi_1, \ldots, \pi_n$ could assure me that κ cannot be empirically falsified. This being so, my attitude towards κ would at least be similar to what it would be were I actually to have a proof of κ. Applying this to the case of the AH argument, we see that though it may not itself be a proof of the 4CT, it might still provide a warrant similar in *a priori* character to that provided by a proof.[20]

But does it provide mathematical knowledge? This is a subtler and more difficult question. I can read about a proof in a newspaper or the announcements section of a journal and learn that a certain theorem (e.g., Fermat's Last Theorem, FLT, for short) has been proved. When the publication and my reading of it are both properly judged to be reliable, this learning can amount to knowledge. I can thus know that FLT is provable, and, supposing I know that the methods of proof used are reliable, I can also know FLT.

I can know all these things—I and many people in fact do—and still not have *mathematical* knowledge of FLT. The reason, roughly, is that in knowing what I know, I don't know the *mathematical reasons* for FLT. Neither do I know how (i.e. by what reasoning) they guarantee it. So, even though I might know that certain propositions (say, the axioms of second-order Peano Arithmetic) are true and that certain inferences are sound, and know that from these propositions and inferences a proof of FLT can be fashioned, I do not thereby gain mathematical knowledge of FLT.

[20] See Peressini [Peressini 2003], Fallis [Fallis 1996] and [Fallis 1997] for related discussions.

This raises a question concerning the AH proof of the 4CT—namely, whether it provides enough insight into the mathematical reasons for the 4CT to give genuinely mathematical knowledge of it—either to those who designed it (i.e. Appel, Haken and Koch) or to others with similar substantial knowledge of the program believed to be implemented by the computing device. I would estimate that it does, but my main point is that the mere fact that at various points the proof is turned over to a computer would not prohibit its being an adequate statement of mathematical reasons for its conclusion.[21]

A key distinction here is that between *program* and *implementation of* program. Knowledge of the computer-assisted proof of the 4CT can give sufficient knowledge of reasons for the 4CT only if two conditions are satisfied. The first is that the knower have extensive enough knowledge of the reducibility program to allow him properly to judge that it is an adequate program for doing what it's intended and/or believed to do. The second is that the knower should have sufficient knowledge of the implementation to warrant judgment that the aforementioned program is indeed the program executed by the machine whose output is used.

It seems the principal novelty of the AH proof concerns the second condition. In ordinary computations, knowledge that a computation implements a program comes through survey of the computation. In the case of the AH proof this is not and perhaps can not be the case. Whether this introduces an empirical element into the AH proof that is fundamentally unlike that which figures in more ordinary proofs may be doubted. This notwithstanding, the question of whether the type of knowledge an informed knower can have of the program that figures in the AH proof and of its implementation can amount to proper knowledge of the *reasons* for the 4CT is an important one, and one that deserves more careful discussion than I can give it here.[22]

4 *Formalization and Rigor*

The prevailing view of proof sees rigor as a necessary feature of proof and formalizability as a necessary condition of rigor. On this view, a rigorous proof is one that can be known not to conceal substantive (i.e. non-logical) information. Its inferences can be seen to be valid solely by appeal to logical relations between concepts and not to their 'senses' or 'contents'. As Pasch put it:

> ... the process of inferring must always be independent of the sense of... concepts just as it must be independent of diagrams. It is only relations between ... concepts that should be taken into account in the propositions and definitions that are dealt with. In the course of the deduction, it is certainly legitimate and useful, though by no means necessary, to think of the reference of the concepts involved. If it is indeed necessary to so think, the defectiveness of the deduction and the inadequacy of the ... proof is thereby revealed unless it is possible to remove the gaps by modification of the reasoning used.
>
> [Pasch 1912], p. 98[23]

---

[21] See Fallis [Fallis 2003] for an interesting discussion of the phenomenon of "gaps" in the statements of "reasons" offered by proofs.

[22] For an indication of one direction such further discussion might take, see Fallis [Fallis 2002] where broad questions concerning the fit between the goals of mathematics and its methods are considered.

[23] My translation. The same basic idea was defended in the middle of the eighteenth century by J. H. Lambert (cf. [Lambert 1766], p. 162).

Rigorous proof, on this view, is reasoning all of whose inferences track purely logical relations between concepts. In the late nineteenth and early twentieth centuries, syntactical criteria for such relations were developed and these have become the basis for the currently prevailing view of formalization.

The reasoning behind this view is straightforward: (i) proper proofs are proofs that either are or can readily be made rigorous; (ii) proofs that are or can readily be made rigorous are formalizable; therefore (iii) all proper proofs are formalizable. Call this argument the *common argument* and its conclusion the *common view*.

Both the view and its argument seem dubious. Mathematical proofs are not commonly formalized, either at the time they're presented or afterwards. Neither are they generally presented in a way that makes their formalizations either apparent or routine. This notwithstanding, they are commonly presented in a way that *does* make their *rigor* clear—if not at the start, then at least by the time they're widely circulated among peers and/or students. There are thus indications that rigor and formalization are independent concerns.

This is not the common view, however. On that view, non-formalized proofs are typically close enough to formalized proofs to make the fact of formalizability clear and the remaining work of formalization routine. Saunders Mac Lane maintained such a view.

> A mathematical proof is rigorous when it is (or could be) written out in the first order predicate language $L(\in)$ as a sequence of inferences from the axioms ZFC, each inference made according to one of the stated rules.... practically no one actually bothers to write out... formal proofs. In practice, a proof is a sketch, in sufficient detail to make possible a routine translation of this sketch into a formal proof.
>
> [Mac Lane 1986], p. 377

This, as I said, is the common view. But common or not, not everyone agrees, and the dissenters include some who have great experience in the work of formalization. The artificial intelligence researcher John A. Robinson (an expert in automated theorem-proving) is a case in point. His experience with formalization causes him to remark that it is often "surprisingly difficult" and only occasionally a routine matter (cf. Robinson [Robinson 1997], p. 54).

> In most cases it requires considerable ingenuity, and has the feel of a fresh and separate mathematical problem in itself. In some cases... formalization is so elusive as to seem to be impossible. [Robinson 1997], p. 54

Still more importantly he sees *standard* formalization—that is, formalization of the usual reduction-to-syntactically-presented-rudimentary-logical-inference variety—as often undesirable even if manageable. The reason is that a prime goal of proof is explanation (cf. [Mac Lane 1986], pp. 378–79 and [Robinson 2000], p. 277) and standard formalization can obscure explanatory connections. Indeed, Robinson believes that it "typically destroys all traces of the explanatory power of the informal proof" (Robinson [Robinson 1997], p. 56; see also Robinson [Robinson 2000], pp. 293–94).

That this is so is due to the fact that standard formalization breaks an informal proof down into many artificially small steps of reasoning. Explanation, on the other hand, is typically carried by "large-scale, high-level" ([Robinson 2000], p. 279) "architectural" patterns of reasoning, patterns which may be obscured when embedded in a mass of rudimentary logical inferences.

> Too much detail causes difficulty in viewing the big picture. One cannot see the forest for the trees. [Robinson 2000], p. 479[24]

Robinson believes that real mathematical proofs are essentially *performances*, and not "structured static texts" ([Robinson 2000], p. 281) in which nothing *happens*. To focus on formalized proofs is to view proofs as such texts, and to do this is like experiencing music only by reading musical scores (cf. *loc. cit.*). Scores are important, but there is more to a 'living' piece of music than its score.

In the same way, there is more to a real proof than its formalization. A formal proof is "only the score, only the script, only the instructions for producing the real proof" (*loc. cit.*). Indeed, it's typically not even that, since it's generally an afterthought rather than a guide to proof.[25]

Robinson supports this view by outlining a form of performative experiment—an introspective experiment in which he looks for theorems that he (a) understands, but which he (b) finds incredible and for which he (c) possesses a proof that is within his power to understand with a reasonable effort. Once a selection is made, the "experiment" consists in learning how to 'perform' the given proof while monitoring the process to detect the "crucial moments in the proof" (*op. cit.*, p. 282) when his attitude turns from incredulity to acceptance. Robinson illustrates these ideas with an example from number theory, namely, Erdős's proof of Bertrand's Conjecture that for every positive integer $n > 1$, there is at least one prime p such that $n < p < 2n$.[26]

The result of such an experiment, Robinson suggests, is that a proof comes to be stored as "a collection of relatively few leading ideas dealing with interesting . . . phenomena" ([Robinson 2000], p. 291). Increasing familiarity with these phenomena eventually gives them "an aura of certainty" and they become "established resources" which can be "triggered at will". The mind, Robinson says, is "hungry" (*op. cit.*, p. 292) for such "key ideas" that capture the gist of a proof. Fixing our attention on them provides a better grasp of the overall plan of a proof and, with the overall plan before it, the mind can then "understand" (*loc. cit.*) the proof and not get lost in its details. Details can thus be the enemy of understanding and blind us to the overall architecture of a proof (cf. Robinson [Robinson 2000], p. 292). The central work to be done, then, is to identify those larger patterns of inference that guide actual mathematical practice and understanding and make a formal protocol (or different local formal protocols) of them.

---

[24] There is an unmistakable parallel here with Poincaré, who campaigned against the "logicization" of mathematical reasoning for similar reasons. See Poincaré [Poincaré 1905], ch. I. See also Detlefsen [Detlefsen 1992] for a fuller discussion of these ideas of Poincaré's.

[25] For related though in certain respects broader discussions of mathematical *activities*, see Giaquinto [Giaquinto 2005a] and Rota [Rota 1997]. In addition to proof, the former identifies discovery, justification and explanation as other key epistemic activities. The latter describes mathematical practice as concerned with such things as investigations, intuitions, conjectures and verifications, all of which are taken to be different from, albeit related to, proof. Rota also discusses the axiomatic method and how it can sometimes conceal explanatory connections, and offers a few observations concerning the processes through which mathematical reasoning is refined.

[26] The conjecture was first proved in 1850 by Chebychev. Erdös's proof is more picturesque than Chebychev's, however, and (therefore?) more memorable, or so Robinson argues (cf. Robinson [Robinson 2000], pp. 283–89).

Robinson notes certain difficulties involved in attempting to do this (cf. Robinson [Robinson 2000], pp. 293–294). There is, though, a larger possibility that Robinson seems to overlook—namely, that there may simply be no family of perceivable entailments that (a) are individually "larger" than those of rudimentary logic and (b) offer adequate protection from the admission of dangerous gaps in mathematical reasoning. This, at any rate, is a central problem confronting Robinson's and similar proposals.

There is also a question concerning a possible deeper relation between explanatory content/character and rigor. Traditionally at least, mathematical reasoning has been taken to be at its most rigorous when it is also at its most potently explanatory. We're most certain to avoid gaps in reasoning when premises *explain* conclusions.

Hilbert suggested such a view when he wrote

> It is an error to believe that rigor in proof is an enemy of simplicity. On the contrary we find it confirmed by numerous examples that the rigorous method is at the same time the simpler and the more easily comprehended. The very effort for rigor forces us to find out simpler methods of proof. [Hilbert 1902], p. 441

In any event, it seems at least possible to think of the rigor as linked to explanatory transparency—an inference being rigorous to the extent that its premises can be seen to *explain* its conclusion.[27] The greater such explanatory transparency, the more confident we can be that unrecognized information has not been used to connect a conclusion to premises in ways that matter. To the extent, then, that formalization decreases explanatory transparency, it also decreases rigor. A reexamination of the commonly presumed connection(s) between rigor and formalization would thus seem to be in order.

5 Visualization and Diagrammatic Reasoning in Mathematics

The common view of diagrams and their role in proof has for some time been that they are merely heuristic devices, useful instruments to aid the discovery, formulation and/or the intuitive comprehension of proofs, but lacking any genuinely justificative role in proof. Leibniz stated the essentials of this view as follows.

> . . . geometers do not derive their proofs from diagrams, though the expository approach makes it seem so. The cogency of demonstration is independent of the diagram, whose only role is to make it easier to understand what is meant and to fix one's attention. It is universal propositions, i.e. definitions and axioms and theorems which have already been demonstrated, that make up the reasoning, and they would sustain it even if there were no diagram. [Leibniz 1981], p. 360[28]

---

[27] For further recent discussions of the role of explanation in mathematics, see Mancosu [Mancosu 2000] and [Mancosu 2001], Tappenden [Tappenden 2005] and Mancosu, Jorgensen *et al.* [Hafner/Mancosu 2005].

[28] The *New Essays* were published posthumously in 1765. It was written over an extended period of time and completed sometime between 1709 and Leibniz' death in 1716.

Many have advocated similar views,[29] although some have disagreed, the preeminent example being Kant. In Kant's view, diagrammatic reasoning (or something like it) was not only to be admitted into genuine proof, it was generally necessary for it ([Kant 1781–87], A716–17/B744–45; A713–14/B741–42.)

Less radical, but still supportive of the use of diagrams in geometrical reasoning were Hobbes ([Hobbes 1655], [Hobbes 1656]), Newton ([Newton 1720], appendix, pp. 229–230), Locke ([Locke 1697], p. 58) and such lesser figures as Francis Maseres ([Maseres 1758], pp. ii–iii). Some, indeed, went beyond sympathy. C. S. Peirce, for example, maintained that virtually all reasoning—logical as well as mathematical—is either diagrammatic overall or has essential diagrammatic aspects ([Peirce 1898]).

Historically, there have been two main reasons for denying a genuinely justificative role to diagrams. One is unreliability (see [Hahn 1933] for a summary statement), the other their *particularity* ([Locke 1689–90], Bk IV, ch 1, sect 9; [Hume 1748], sect XII, part I; [Berkeley 1709], in [Berkeley 1948–57], vol. 1, p. 221, [Berkeley 1710], Bk IV, ch 7, sects 7–13). In geometry, there are well-known examples used to support the charge of unreliability. A widely used example is the famous diagrammatic "proof" that all triangles are equilateral. This was a favorite of Hilbert's which he repeated in various of his lecture courses on the foundations of geometry.

The other, perhaps more fundamental reason for denying justificative status to diagrammatic reasoning is their *particularity*. Mathematical truths are typically general truths while diagrams are *particular* figures. Since deductive reasoning concerning a particular figure can not establish a general truth, diagrammatic reasoning can not deductively justify a typical mathematical truth. It can only do so by some sort of analogical or broadly inductive extension. So, at any rate, the traditional reasoning goes.

In recent decades there has been renewed interest in diagrammatic reasoning in logic and mathematics. One influential example is the investigation and defense initiated by Jon Barwise and John Etchemendy, and pursued by various of their students and others. Barwise and Etchemendy argue that diagrams can and often do play a genuine epistemic role in proof: "we claim that visual forms of representation can be important, not just as heuristic and pedagogic tools, but as legitimate elements of mathematical proofs" ([Barwise/Etchemendy 1991], p. 9). Later they strengthen this by saying that "diagrams and other forms of visual representation can be *essential* and legitimate components in valid deductive reasoning" (*op. cit.*, p. 16, emphasis added).

They offer two responses to the charge that diagrammatic (or broadly visual) reasoning is unreliable. The first is the basic logical point that the existence of fallacious instances of reasoning of a given broad type does not impugn *all* instances of that type. Accordingly, even though some diagrammatic reasoning is fallacious, not all of it need be.

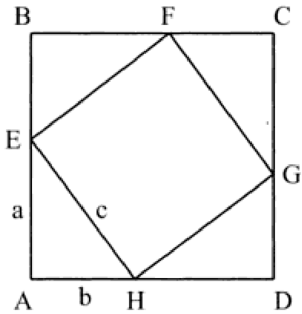
The second is a set of specific examples of what Barwise and Etchemendy characterize as 'perfectly valid proofs' that use diagrammatic (or other visual) reasoning in justificative ways ([Barwise/Etchemendy 1991], p. 12). Their mathematical example is an argument for

---

[29] We quoted a remark from Pasch earlier that expressed such a view. Hilbert too said such things (cf. lecture notes on geometry of the summer semester of 1894, the winter semester of 1898/99 and the summer semester of 1927). It is more difficult to determine his final view, however, because of the emphasis he elsewhere placed on 'intuitive grasp' (*anschaulichen Erfassen*) in geometric thinking ([Hilbert/Cohn-Vossen 1932], V, VI).

Russell too held such views ([Russell 1901], pp. 88–89, [Russell 1919], p. 145), as did Hans Hahn ([Hahn 1933]). More recent examples include Dieudonné [Dieudonné 1960], p. v and Tennant [Tennant 1986], p. 304.

the Pythagorean theorem ([Barwise/Etchemendy 1991], p. 12–13, Example 3) that combines diagrammatic reasoning and algebraic reasoning. The basic diagram is as follows:



The focal triangle is $\triangle EAH$, and the claim to be proved is that $a^2 + b^2 = c^2$. The argument begins by constructing a square on EH and replicating $\triangle EAH$ three times as indicated in the diagram. One then determines that each side of $ABCD$ is a straight line by appealing to the theorem that the sum of the angles of a triangle is a straight line. From this, we're told, "one easily sees" that $ABCD$ is itself a square, and a square whose area can be computed in two ways: $(a+b)^2$ and $c^2 + 4(\frac{ab}{2})$. Equating these two and doing the obvious calculations, we arrive at $a^2 + b^2 = c^2$, which is what was to be proved.

Commenting on this argument, Barwise and Etchemendy ([Barwise/Etchemendy 1991], p. 12) make five claims.

1. It is (clearly) a "legitimate proof of the Pythagorean theorem."
2. It is a "combination of geometric manipulation of a diagram and algebraic manipulation of nondiagrammatic symbols."
3. The diagrammatic elements "play a crucial role in the proof."
4. The diagrammatic elements are primary (and typical of many traditional geometric proofs) in two related ways.
   (a) They make the algebraic steps of the argument "almost transparent." Once the diagrammatic steps are in place, the algebraic steps are easy to devise.
   (b) An analogous 'linguistic' proof would be both difficult to discover and difficult to remember without the use of diagrams. (This may give a sense for the suggestion noted above that the diagrammatic elements of the argument are "essential" to it.)
5. The proof (clearly) does not make use of "accidental features" of the diagrams involved.

Barwise and Etchemendy do not argue for these claims and, to my mind, none are evident. I'll briefly state some of my reservations below. Before doing that, though, I'd like to mention that both 4(a) and 4(b) are consonant with the traditional view of diagrams—namely, that they can be heuristically valuable, even though they play no legitimate justificative role in proof. Accepting 4(a) and 4(b) would thus not commit one to a justificative role for diagrams.

Regarding 1, I note two points. First, since the argument does not clearly identify all the different propositions it depends upon, it seems wrong to say that it is clearly a legitimate proof. In fact, it contains gaps, which legitimate proofs are not supposed to have. As an example of a gap, consider the step where Barwise and Etchemendy call for the "replication" of the original triangle three times as "shown" in the diagram. What goes into such "replication"? And what justifies it? There are plausible answers to these questions, but they need to be added to the argument if it is to deserve the title of 'proof'.

Nor does the Barwise-Etchemendy argument seem to support the suggestion in 4(a). In particular, its diagrammatic starting point does not seem to determine its algebraic details or make them 'transparent'. Indeed, as Yanney and Calderhead ([Yanney/Calderhead 1898]) argued long ago, there are at least four significantly different ways to carry a proof of the Pythagorean Theorem forward from the initial diagrammatic starting point of constructing a square on the hypotenuse of $\triangle EAH$. There may be (and perhaps typically is) some point in the development of a diagrammatic argument where a set of algebraic steps capable of completing the argument becomes "transparent". This would not be enough to show, though, that it's the diagrams that produce the transparency. It might instead be sheer accumulation of information, be it diagrammatically or non-diagrammatically supplied. Is this what is happening in the Barwise-Etchemendy proof? Or is there some distinctive gain in transparency that is specifically due to their use of diagrams?

Regarding claim 2, my main concern is clarity: specifically, the clarity of the key notion of 'geometrical manipulation'. On some level, the claim is uncontroversial. There are surely elements of the Barwise-Etchemendy argument that in some sense(s) are *geometrical manipulations*. A more difficult question, though, is what role specifically *visual* information plays in these operations. The answer to this is not, I think, clear, and this lack of clarity mounts when we consider claim 2 in conjunction with claims 3 and 5.

What are the 'geometrical manipulations' that supposedly both reflect essentially visual information and are clearly not accidential? It's hard to see what they might be. The likely candidate is the so-called 'replication' of the original triangle on the four sides of the square constructed on its hypotenuse. But it should be noted that this cannot be regarded as non-accidental unless we're able to establish non-accidentally that the sides of $ABCD$ are straight lines.

Barwise and Etchemendy rightly recognize that there needs to be a proof that the sides of $ABCD$ are straight lines. What is not clear from their argument is that this relies in any significant way on the visual information in their diagrams. Their reasoning is basically that the sides of $ABCD$ are straight lines because the interior angles of a triangle sum to a straight line. This is non-accidentally true of the sides of $ABCD$, however, only because of the similarity (in the geometrical sense) of triangles AEH, BFE, CGF and DHG. It does not derive from the visual qualities of their diagram. We know that AB, for example, is a straight line because: (i) the sum of the interior angles of a triangle is a straight angle and (ii) $\angle AEB$ is equal to such a sum. That (ii) is non-accidentally true follows from the additional facts that (a) $\angle AEB$ is composed of $\angle HEA$, $\angle HEF$ and $\angle FEB$, that (b) $\triangle AEH$ is composed of $\angle HAE$, $\angle AHE$ and $\angle HEA$, and that (c) $\angle HEA = \angle HEA$, $\angle HEF = \angle HAE$ and $\angle FEB = \angle AHE$. Similarly for the other sides of $ABCD$.

Where is the appeal to visual information in all of this? Nothing suggests that it's in (i). We are thus left with (ii). But where in (ii)? Not from (b), since that comes from the definition of $\triangle AEH$. Not from (c) either, though, since we get that from logic ($\angle HEA = \angle HEA$) and the knowledge that $\triangle AEH$ and $\triangle BFE$ are similar. This latter knowledge is not genuinely visual since it comes from knowledge that $\angle HAE$ and $\angle EBF$ are both right angles (because of the way $ABCD$ is constructed), that the angles in a triangle sum to two right angles ((i) again), that $\angle HEF$ is a right angle (due to $EFGH$'s being constructed as a square), and that $\angle HEA$ and $\angle FEB$ are therefore complementary. From this it follows that $\angle FEB = \angle AHE$ and $\angle AEH = \angle BFE$.

There are, of course, various heuristic roles that visual information may play in this reasoning—for example, in first *suggesting* that $\angle AEB = \angle AEH + \angle HEF + \angle FEB$. What is not so clear, however, is what justificative role this information might play.

Suppose, for the sake of argument, that knowledge that $\angle AEB = \angle AEH + \angle HEF + \angle FEB$ *does* come from the visual experience of the diagram. The *content* of that experience—$\angle AEB = \angle AEH + \angle HEF + \angle FEB$—plays a logical role in the argument. This notwithstanding, this content, and its logical relations to the contents of the other elements of the argument, are not the *only* things that affect the type of warrant the argument provides for its conclusion.

Traditionally, proofs have been intended to support belief in the necessity of their conclusions. Basing belief that $\angle AEB = \angle AEH + \angle HEF + \angle FEB$ (or any other premise of the Barwise-Etchemendy argument) on visual information would not provide for the realization of this intention. If our only reason for believing that $\angle AEB = \angle AEH + \angle HEF + \angle FEB$ is that it visually appears to be the case, we will not be in a position at the end of the Barwise-Etchemendy argument to know that it's necessarily the case that $a^2 + b^2 = c^2$. This, it seems to me, is similar to Berkeley's and Hume's objections to Locke's view of diagrams.

There are other questions and concerns raised by the Barwise-Etchemendy proposal, but in the space that remains I'll consider other recent work concerning the use of diagrams and visual information in mathematics. One example is Norman ([Norman 2006]), which argues for a broadly Kantian viewpoint according to which diagrammatic reasoning of the type found in classical geometry can contribute to *a priori* justification. That this is so is due to the fact that the reasoner typically forms concepts of (types of) geometrical objects, that she reasons with diagrams by taking them to represent instances of these concepts and that she then infers a general conclusion by taking the diagram to represent not merely a particular picture or image itself but the full set of such images producible by the same essential process of construction by which the given diagram was constructed. Having argued this, Norman nonetheless concedes that, in the end, diagrammatic reasoning lacks the rigor generally required of proofs. He thus suggests, in the end, that proof is but one means of attaining mathematical knowledge, an idea not unlike that suggested by Wallis, Tymoczko and others.

James Robert Brown has defended a similar view (cf. Brown [Brown 1999], pp. 24–43), arguing that diagrams and pictures can provide evidential grounds for propositions concerning mathematical objects we do not see. Since, however, the propositions supported by diagrams are often more general than the diagrams themselves are, the justificative role of diagrams can not generally be due to their *depiction* of the subject-matters of the propositions they support. They're not so much pictures, says Brown, as 'windows into Plato's heaven' ([Brown 1997], p. 174).

What these 'windows' are and how they're supposed to work is not clear. Brown mentions a "structural similarity" ([Brown 1997], p. 173) between diagrams and what they depict, and maintains that this somehow unites the items that belong to the justificative range of a diagram. This structural similarity is, however, presumably different from what Barwise and Etchemendy had in mind when they claimed that a 'good diagram is isomorphic, or at least homomorphic, to the situation it represents' ([Barwise/Etchemendy 1991], p. 22). Brown at any rate emphasizes that diagrams are not generally either isomorphic or homomorphic to what they represent ([Brown 1997], p. 173).

In the end, Brown doesn't show what he claims to show—namely, that diagrams either constitute proofs or play justificative roles in them. Indeed, some of what he says goes against this. He notes, in particular, that the main epistemic function of diagrammatic reasoning is to provide rational conviction rather than understanding (cf. Brown [Brown 1999], pp. 42–43), the latter being typically reserved for the more conventional 'propositional' proofs. But one must then wonder whether diagrammatic reasoning supports the higher forms of mathematical knowledge. Brown has nothing convincing to say on this point.

Marcus Giaquinto offers yet another account of visual reasoning in [Giaquinto 1994] and [Giaquinto 2005]. Where Brown emphasizes the legitimate evidential force of visual reasoning in establishing various theorems of number theory and analysis (e.g. Bolzano's Intermediate Value Theorem), Giaquinto emphasizes the differences between geometry and analysis, and argues for a much less extensive role for visual reasoning in analysis than in geometry.

Giaquinto's account of the epistemic role of visual reasoning also differs from Brown's. He maintains that visual reasoning is a legitimate means of discovery, where 'discovery' for him has a justificative aspect.[30] "One *discovers* a truth (which one does not already believe) by coming to believe it *independently* in an *epistemically acceptable* way." ([Giaquinto 1994], p. 790, emphases added). By requiring independence he means to rule out mere reliance on testimony. By *epistemically acceptable* ways he means reliable ways that are not undermined by an agent's other beliefs.

Giaquinto expressly denies that visual reasoning can be used as evidence for various theorems of analysis (e.g. the Intermediate Value Theorem, cf. Giaquinto [Giaquinto 1994], p. 793). He denies, in particular, that it can serve as a legitimate means of discovery, in the sense of the term described above. The reason is that the theorems mentioned exhibit a type of generality that defies discovery (in the above sense) by visual reasoning. In visual reasoning in geometry, generality is achieved because the visual reasoning typically 'brings to mind' a reliable general '*form* of thinking'. In the analytic cases mentioned, similarly general reasoning is unreliable.

In the case of Bolzano's Intermediate Value Theorem, he argues, in particular, for the falsity of the following:

(i) Any continuous function that changes signs on an interval has an uninterrupted curve from a point above the x-axis to a point below it.
(ii) Any function whose curve meets the x-axis has a zero value.

The latter is false because, judged according to visual criteria, a curve with a single point gap at the x-axis will nonetheless *look* like it intersects the x-axis. The two parts of such a line "could not appear to be separated by just one point, as a point has zero breadth" (*op. cit.*, p. 800). Visualization therefore cannot be a way of discovering (in Giaquinto's sense) Bolzano's theorem.

Assumption (i) seems even less defensible. The assumption that every continuous function has an uninterrupted curve is false because not every continuous function has any (visualizable) curve at all. Giaquinto offers Weierstrass' everywhere continuous but nowhere differentiable function as an example. At every stage of visualization (assuming the stages to follow the imagined steps of magnification) of this function there is a smooth part and the function is thus

[30] He distinguishes two different types of justification, however. One (*demonstrative* justification) requires both the absence of any violation of basic standards of rationality and an ability to explicitly give a reason for one's belief. The other (*default* justification) requires only the former (cf. [Giaquinto 1994], p. 791).

differentiable. On the other hand, a curve which is non-differentiable at a point "makes a sharp turn at that point, and a curve consisting of sharp turns at every point, without any smooth segments between sharp points, is unvisualizable" (*op. cit.*, p. 801). Some continuous functions thus have no curves (i.e. no visualizable curves) at all. This being so, theorems pertaining to such functions cannot be discovered through visualization. Or so Giaquinto reasons.

He also discusses what he sees as a signal difference between visual reasoning in geometry and visual reasoning in analysis (*op. cit.*, pp. 804–805). Certain geometrical concepts (e.g. those of a circle and a straight line) can have visual representations because some physical (or visually imagined) figures can appear to be *perfect* exemplars of their geometrical type. Thus, some physical or imagined circles can appear to be *perfectly* circular and some physical or imagined straight lines can appear to be *perfectly* straight. Geometrical concepts thus amount to idealizations or perfections, and can be visually represented by exemplars that are near enough to being perfect that their defects are not visually detectable.

The same is not true of such analytic concepts as continuous function, differentiable function and the integral. The first cannot be visualized by an uninterrupted curve since this both excludes some continuous functions and includes come non-continuous ones. The second cannot be visualized as a function with a smooth curve. The third cannot be visualized because of demands that analysis places on the concept of area. Hence, there are significant differences between geometry and analysis as regards the discoverability of theorems via visualization. Visualization may often be an aid to understanding and a stimulus or "trigger" to discovery in analysis, but it is only rarely a *mode* of discovery (cf. *op. cit.*, p. 811). This notwithstanding, it may still be a valuable tool to the analyst, and one whose value can be expected to grow with increased experience in analysis (*op. cit.*, p. 812).[31]

Thus far, I've not questioned the traditional assumption that diagrammatic reasoning is useful. I'll now briefly consider this question and the growing body of literature concerned with it. It contains some of the most interesting recent work concerning diagrammatic reasoning.

An important earlier study was Jill Larkin & Herbert Simon's "Why a Diagram is (Sometimes) Worth Ten Thousand Words" ([Larkin/Simon 1987]). A key difference between diagrammatic and linguistic (what they and others term 'sentential') reasoning, they claimed, is the degree to which information explicit in the one is implicit in the other. Diagrams characteristically display information explicitly that is only implicit in their linguistic counterparts. Since implicit information has to be computed in order to be used, linguistic reasoning typically involves more computation than diagrammatic reasoning, and this means that it's less easy.[32] Larkin and Simon argue that this is due in large part to the fact that linguistic representation is sequential or linear, while diagrammatic representation is planar. In these planar representations, spatially adjacent parts of the diagram often carry inferentially adjacent information. A process of diagrammatic reasoning is thus commonly driven by visual traversal or survey of a diagram and requires relatively little extraction of (i.e., search for) tacit elements.

[31] For more on the idea that diagrams and/or other types of pictures can act causally as "triggers" for belief-formation see Giaquinto [Giaquinto 2005].

[32] This assumes, of course, that we're talking about linguistic and diagrammatic expressions of the same information. Representations are treated as informationally equivalent when the information in each is inferrable from the information in the other. Informationally equivalent representations are then said to be computationally equivalent when, roughly, every inference in the one is as easy as the parallel inference in the other.

A different explanation of the relative efficiency of diagrammatic over linguistic reasoning is pursued by Stenning and Lemon ([Stenning/Lemon 2001]). They argue that it is typically due to diagrams' having a lower capacity for expression, in particular, a lower capacity for expressing abstractions. They argue further that restricted capacity to express abstractions generally makes for tractability of inference, while enhanced such capacity makes for intractability. The authors broadly attribute these differences to the planar character of diagrammatic representations.

> ... the expressive restrictions on DRs [diagrammatic representations] arise from an interaction between topological and geometrical constraints on plane surfaces, and the ways in which diagrams are interpreted.
>
> [Stenning/Lemon 2001], p. 30, brackets added

The topological constraints mentioned stem from a theorem of Helly's which limits the number of convex regions the mutual inclusion/exclusion relationships and emptiness/non-emptiness features that can be accurately presented in a planar array (cf. Stenning and Lemon [Stenning/Lemon 2001], pp. 45–46). Diagrams that exceed this limit are not generally trustworthy as regards the inclusion/exclusion, emptiness/non-emptiness information they convey.

Stenning and Lemon offer general characterizations of diagrammatic represensentation systems and efficacious diagrammatic representation systems. Roughly, a representation system functions diagrammatically to the extent that its interpretation can be directly read off its spatial characteristics. A little more exactly, a diagrammatic representation is "a plane structure in which representing tokens are objects whose mutual spatial and graphical relations are directly interpreted as relations in the target structure" (*op. cit.*, p. 36).

The directness mentioned plays a key role in the efficiency of diagrammatic reasoning. As the interpretation of a representational system grows in abstractness (i.e., becomes less direct) its diagrammatic character decreases and the need for *extractive* interpretation (hence complexity) increases. Roughly, then, what makes diagrammatic reasoning efficient, when it is efficient, is the *directness* of its interpretation—the relatively great capacity a user has to read off key features of the target structure from the appearance of the diagram. It becomes useful to the extent that the features of the target system that can be directly read off the diagram comprise important features of the target system. See *op. cit.*, pp. 47–48 for a general characterization of diagrammatic effectiveness.

Unfortunately, the examples treated in this paper, as in most other recent work on diagrams, deal mainly with the use of diagrams in purely logical reasoning. Little attention is given to more complicated cases such as the use of diagrams in geometrical reasoning.[33] It may be that the general characterizations of diagrammatic reasoning and effective diagrammatic reasoning that Stenning and Lemon offer can be extended to such cases, but they offer little to support such a view. Nor, finally, do they engage the question of the general relationship between diagrammatic reasoning and proof.

These limitations notwithstanding, I commend the work for its attempt to provide a psychologically plausible explanation of why diagrammatic reasoning seems so often useful. Any serious account of the role of diagrams in proof will ultimately have to come to grips with the

[33] For examples of what mathematicians count as diagrammatic reasoning, see the continuing series of *Proofs without Words* in the *Mathematics Magazine* and also the two books by Nelsen ([Nelsen 1997], [Nelsen 2001]).

issues these authors address. I might also mention that the references in the paper provide the interested reader with valuable suggestions for continued study of these questions.

6 Concluding Thoughts

I've focused on three preoccupations of recent writings on proof:

1. **The role and possible effects of empirical reasoning in mathematics.** Do recent developments (specifically, the computer-assisted proof of the 4CT) point to something essentially new as regards the need for and/or effects of using broadly empirical and inductive reasoning in mathematics? In particular, should we see such things as the computer-assisted proof of the 4CT as pointing to the existence of mathematical truths of which we cannot have *a priori* knowledge?
2. **The role of formalization in proof.** What are the patterns of inference according to which mathematical reasoning naturally proceeds? Are they of 'local' character (i.e. sensitive to the subject-matter of the reasoning concerned) or 'global' character (i.e. invariant across all subject-matters)? Finally, what if any relationship is there (a) between the patterns of inference manifest in a proof and its explanatory capacity and (b) between explanatory capacity and rigor?
3. **Diagrams and their role in mathematical reasoning.** What essentially *is* diagrammatic reasoning, and what is the nature and basis of its usefulness? Can it play a justificative role in the development of mathematical knowledge and, more particularly, in genuine proof? Finally, does the use of diagrammatic reasoning force an adjustment either in our conception of rigor or in our view of its importance?

Concerning 1, I've urged caution as regards the suggestion by Tymoczko (and others) that the computer-assisted proof of the 4CT calls for fundamental changes in our understanding of mathematical method and proof. Its chief novelty, in my view, is the adjustment it suggests in our views of how we may come to know that proofs exist. It offers a concrete illustration of a proof that may defy human surveyal but nonetheless admits of survey by a computational routine designed and verified by humans.

The broader proposal of Jaffe and Quinn to "institutionalize" the use of empirical methods in mathematics does not challenge our understanding of the nature of proof so much as our use of it as a justificative standard in mathematics. It joins questions regarding proper method in mathematics to larger questions of morality and social practice.

The questions raised in 2 remain largely open. Robinson's work emphasizes the importance of finding the patterns that carry the flow of information in mathematical proof, and presents reasons for thinking they're often determined by 'local' topic and are not of a topic-neutral logical character.

The questions raised in 3 remain similarly open. This notwithstanding, insightful cases have been made for the significance of diagrammatic reasoning as justificative (as distinct from purely heuristic). At the same time, our understanding of possible limits on justificative uses of diagrammatic reasoning have been similarly advanced.

As regards the broad questions identified at the beginning of this paper, I've argued that little has been done to challenge the traditional view that proof has a distinctive role to play in the development of mathematical knowledge. In particular, I've argued that there is nothing new in the view that broadly empirical methods can play a role in mathematical investigation.

The challenges by Robinson and others to the traditional view of formalizability as an ideal of proof are of greater interest. They suggest that the level of detail required by certain types of formalization may actually interfere with the recognition of larger-scale structures in proofs upon which their explanatory potential depends.

Finally, the growing body of work on diagrammatic reasoning is of similarly great interest and potential. It challenges traditional ideas concerning the role of diagrammatic reasoning in proof and in the development of mathematical knowledge more generally. It suggests, in particular, that diagrammatic reasoning has a justificative and not merely a heuristic role to play in proof. Much interesting work has already been done in this direction, and more is sure to follow.

References and Bibliography

[Appel/Haken 1977] Appel, K., and Haken, W., "Every planar map is four colorable. I. Discharging", *Illinois Journal of Mathematics* 21 (1977), pp. 429–490.

[Appel et al. 1977] Appel, K., Haken, W. & J. Koch, "Every planar map is four colorable. II. Reducibility", *Illinois Journal of Mathematics* 21 (1977), pp. 1–251.

[Archimedes] Archimedes, *The Method*, preface. Heiberg ed. text with an English translation in [Thomas 2000], pp. 221–223.

[Aristotle 1908] Aristotle, *Posterior Analytics*, Oxford Translations, W. D. Ross and J. A. Smith eds., Oxford University Press, Oxford, 1908–1954.

[Aristotle 2000] ——, *Nichomachean Ethics*, English trans. and ed. by Roger Crisp, Cambridge University Press, Cambridge &New York, 2000.

[Barwise/Etchemendy 1991] Barwise, J. and J. Etchemendy, "Visual information and valid reasoning", in *Visualization in teaching and learning mathematics*, W. Zimmermann & S. Cunningham eds., Mathematical Association of America, Washington, DC, 1991.

[Berkeley 1948–57] Berkeley, G., *The Works of George Berkeley, Bishop of Cloyne.*

[Berkeley 1709] ——, *An Essay Towards A New Theory Of Vision*, in [Berkeley 1948–57], vol. 1.

[Berkeley 1710] ——, *A Treatise concerning the Principles of Human Knowledge*, in [Berkeley 1948–57], vol. 2.

[Borwein 1992] Borwein, J., "Some observations on computer assisted analysis", *Notices of the American Mathematical Society* 39 (1992), pp. 825–829.

[Brown 1997] Brown, J. R., "Pictures and Proofs", *British Journal for the Philosophy of Science* 48 (1997), pp. 161–80.

[Brown 1999] ——, *Philosophy of Mathematics:An introduction to the world of proofs and pictures*, Routledge, London & New York, 1999.

[Cicero 1894] Cicero, M. T., *The orations of Marcus Tullius Cicero*, 4 vols., London, G. Bell & sons, 1894–1903.

[Descartes 1620–28] Descartes, R., "Rules for the Direction of the Mind", in *The Philosophical Writings of Descartes*, vol. I, trans. by J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge University Press, Cambridge, England, 1985.

[Descartes 1637] ——, *La Géométrie*, 1637. First published as an appendix to his *Discourse on Method.* Page references are to the English translation by D. Smith and L. Latham, Dover, New York, 1954.

[Detlefsen/Luker 1980] Detlefsen, M. & M. Luker, "The Four-Color theorem and mathematical proof", *Journal of Philosophy* 77 (1980), pp. 803–820.

[Detlefsen 1992] ——, "Poincaré Against the Logicians", *Synthese* 90 (1992), pp. 349–378.

[Detlefsen 2004] ——, "Formalism", ch. 8 of *The Oxford Handbook of Philosophy of Mathematics and Logic*, S. Shapiro (ed.), Oxford University Press, Oxford, 2004.

[Dieudonné 1960] Dieudonné, J., *Foundations of Modern Analysis*, Academic Press, New York, 1960.

[Euclid 1956] Euclid, *The Elements*, English trans. by Sir Thomas Heath, second revised ed., 3 vols., Dover, New York, 1956.

[Ewald 1996] Ewald, W., *From Kant to Hilbert: A source book in the foundations of mathematics*, two volumes, W. Ewald (ed.), Oxford University Press, Oxford, New York, etc., 1996.

[Fallis 1996] Fallis, D., "Mathematical proof and the reliability of DNA evidence", *American Mathematical Monthly* 103 (1996), pp. 491–497.

[Fallis 1997] ——, "The Epistemic Status of Probabilistic Proof", *Journal of Philosophy* 94 (1997), pp. 165–186.

[Fallis 2002] ——, "What Do Mathematicians Want?: Probabilistic Proofs and the Epistemic Goals of Mathematicians", *Logique et Analyse* 45 (2002), pp. 373–388.

[Fallis 2003] ——, "Intentional gaps in mathematical proofs", *Synthese* 134 (2003), pp. 45–69.

[Giaquinto 1992] Giaquinto, M., "Visualizing as a Means of Geometrical Discovery", *Mind and Language* 7 (1992), pp. 382–401.

[Giaquinto 1993] ——, "Visualizing in Arithmetic", *Philosophy and Phenomenological Research* 53 (1993), pp. 385–396.

[Giaquinto 1994] ——, "Epistemology of Visual Thinking in Elementary Real Analysis", *British Journal for the Philosophy of Science* 45 (1994), pp. 789–813.

[Giaquinto 2005] ——, "From Symmetry Perception to Basic Geometry", in [Mancosu et al. 2005]

[Giaquinto 2005a] ——, "Mathematical Activity", in [Mancosu et al. 2005].

[Hafner/Mancosu 2005] Hafner, J. and P. Mancosu, "The Varieties of Mathematical Explanation", in [Mancosu et al. 2005].

[Hahn 1933] Hahn, H., "The Crisis in Intuition", in H. Hahn *Empiricism, Logic, and Mathematics: Philosophical Papers*, B. McGuinness ed., D. Reidel Publishing Co., Dordrecht, 1980.

[Hilbert 1902] Hilbert, D., "Mathematical Problems", *Bulletin of the American Mathematical Society* **8** (1902), pp. 437–479. English translation of "Mathematische Probleme", *Archiv der Mathematik und Physik* (3rd series) **1** (1901), pp. 44–63, 213–237. The latter is a transcription of Hilbert's 1900 address to the World Congress of Mathematicians.

[Hilbert/Cohn-Vossen 1932] Hilbert, D. & S. Cohn-Vossen, *Anschauliche Geometrie*, Julius Springer, Berlin, 1932.

[Hobbes 1655] Hobbes, T., *De Corpore*, English trans. in [Hobbes 1839–45], vol. 1.

[Hobbes 1656] ——, *Six Lessons to the Professors of the Mathematics, one of Geometry, the other of Astronomy, in the chairs set up by the noble and learned Sir Henry Savile, in the University of Oxford*, in [Hobbes 1839–45], vol. 7.

[Hobbes 1839–45] ——, *The English works of Thomas Hobbes of Malmesbury*. W. Molesworth (ed.). Eleven volumes. J. Bohn, London, 1839–1845.

[Hume 1748] Hume, D., *An Enquiry concerning Human Understanding*, first published as *Philosophical Essays concerning Human Understanding*, London. Reprinted as *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge and P.H. Nidditch, Clarendon Press, Oxford, 1978.

[Hutton 1795] Hutton, C., *A Mathematical and Philosophical Dictionary*, J. Johnson, and G.G. and J. Robinson, London, 1795–96. Two volumes. Reprinted by G. Olms Verlag, Hildesheim and New York, 1973. Reprinted in four volumes by Thoemmes Press, Bristol, 2000.

[Jaffe/Quinn 1993] Jaffe, A. & F. Quinn, "'Theoretical mathematics': Towards a cultural synthesis of mathematics and theoretical physics", *Bulletin of the American Mathematical Society* 29 (1993), pp. 1–13.

[Jesseph 1999] Jesseph, D., *Squaring the circle: The War between Hobbes and Wallis*, University of Chicago Press, Chicago, 1999.

[Kant 1781–87] Kant, I., *Critique of Pure Reason*, English trans. and ed. by P. Guyer and A. Wood, Cambridge University Press, Cambridge, 1998. The first edition of this work appeared in 1781, the second in 1787.

[Lambert 1766] Lambert, J., "Theorie der Parallelinien". This was written in 1766 but not published until 1786, when J. Bernoulli prepared an edition of Lambert's work. It is reprinted in *Die Theorie der Parallelinien von Euklid bis auf Gauss*, F. Engel and P. St'ackel eds., Teubner, Leipzig, 1895. Page references are to this reprinting.

[Larkin/Simon 1987] Larkin, J. and H. Simon, "Why a Diagram is (Sometimes) Worth Ten Thousand Words", *Cognitive Science* 11 (1987), pp. 65–100.

[Leibniz 1707] Leibniz, G. W. F., *Opera philosophica quae exstant latina, gallica, germanica omnia*, J. E. Erdmann (ed.), Scientia, Aalen, 1959.

[Leibniz 1981] ——, *New Essays Concerning Human Understanding*, P. Remnant and J. Bennett trans., Cambridge University Press, Cambridge, 1981.

[Locke 1689–90] Locke, J., *An Essay concerning Human Understanding*, P. H. Nidditch ed., Oxford University Press, Oxford, 1975.

[Locke 1697] ——, "Letter to Edward, the Bishop of Worcester", vol. 4 *Philosophical Works & Selected Correspondence of John Locke*, InteLex, Charlottesville, 1997.

[Mac Lane 1986] Mac Lane, S., *Mathematics: Form and Function*, Springer, New York, 1986.

[MacLaurin 1742] MacLaurin, C., *A treatise of fluxions*, T.W. & T. Ruddimans, Edinburgh, 1742. Excerpts reprinted in [Ewald 1996], volume 1. Imprint New York, Johnson Reprint Corp., 1964-67. Page references are to this reprinting.

[Mancosu 2000] Mancosu, P., "On Mathematical Explanation", in *The growth of mathematical knowledge*, E. Grosholz and H. Breger eds., Synthese Library volume 289, Kluwer, Dordrecht, 2000.

[Mancosu 2001] ——, "Mathematical Explanation: Problems and Prospects", *Topoi* 20 (2001), pp. 97–117.

[Mancosu et al. 2005] Mancosu *et al.*, *Visualization, Explanation and Reasoning Styles in Mathematics*, P. Mancosu, Klaus Frovin Jørgensen & Stig Andur Pedersen eds., Synthese Library, volume 327, Springer, Dordrecht, 2005.

[Mancosu 2005a] Mancosu, P., "Visualization in Logic and Mathematics" in [Mancosu et al. 2005].

[Marx 1844] Marx, K., *Comments on James Mill,* Éléments D'économie Politique. English trans. by C. Dutt in *Karl Marx, Frederick Engels: Collected Works*, volume 3. International Publishers, New York, 1975.

[Maseres 1758] Maseres, F., *A dissertation on the use of the negative sign in algebra: containing a demonstration of the rules usually given concerning it*, Printed, by S. Richardson, London, 1758.

[Nelsen 1997] Nelsen, R., *Proofs Without Words: Exercises in Visual Thinking*, Math. Assoc. Amer., Washington, DC, 1997.

[Nelsen 2001] ——, *Proofs Without Words II: More Exercises in Visual Thinking*, Math. Assoc. Amer., Washington, DC, 2001.

[Newton 1720] Newton, I., *Universal Arithmetick, or a Treatise of Arithmetical Composition and Resolution*. Translated from the Latin by J. Raphson. London, 1720.

[Norman 2006] Norman, J., *After Euclid: Visual Reasoning & the Epistemology of Diagrams*, CSLI, Stanford, 2006.

[Pappus 2000] Pappus, *The Treasury of Analysis*, in [Thomas 2000].

[Pasch 1912] Pasch, M., *Vorlesungen über neuere Geometrie*, 2nd, ed., Teubner, Leipzig, 1912. First ed. 1882.

[Peirce 1898] Peirce, C. S., "The Logic of Mathematics in Relation to Education", *Educational Review* 8 (1898), pp. 209–216. Reprinted in [Ewald 1996], vol. I.

[Peressini 2003] Peressini, A., "Proof, Reliability, and Mathematical Knowledge", *Theoria* 69 (2003), pp. 211–232.

[Poincaré 1902] Poincaré, H., *Science and Hypothesis*, English trans. of 1905 French original. Dover, New York, 1952.

[Poincaré 1905] ——, *The Value of Science*, English trans. by G. B. Halsted of 1905 French original. Dover, New York, 1958.

[Robinson 1997] Robinson, J., "Informal Rigor and Mathematical Understanding", in *Computational Logic and Proof Theory*. Proceedings of the 5th annual Kurt Gödel Colloquium, August 25–29, 1997. Springer, Heidelberg & New York, 1997.

[Robinson 2000] ——, "Proof = Guarantee + Explanation", in S. Hölldobler (ed.), *Intellectics and computational logic: papers in honor of Wolfgang Bibel*, Kluwer Academic Publishers, Dordrecht (the Netherlands) and Boston, 2000.

[Robinson 2004] ——, "Logic is not the whole story", in *First-order logic revisited*, V. Hendricks et al. (eds.), Logos, Berlin, 2004.

[Rota 1997] Rota, G.-C., "The Phenomenology of Mathematical Proof", in *Proof and Progress in Mathematics*, A. Kanamori ed., special issue, *Synthese* 111(2) (1997), pp. 183–196.

[Russell 1901] Russell, B., "Mathematicians and Metaphysicians", in *Mysticism and Logic*, Doubleday Anchor, 1957.

[Russell 1919] ——, *Introduction to Mathematical Philosophy*, George Allen and Unwin Ltd., London, 1919.

[Smith 1776] Smith, A., *An inquiry into the nature and causes of the wealth of nations*, printed for W. Strahan and T. Cadell, London, 1776.

[Stenning/Lemon 2001] Stenning, K. & O. Lemon, "Aligning logical and psychological perspectives on diagrammatic reasoning", *Artificial Intelligence Review* 15 (2001), pp. 29–62.

[Sturm 1700] Sturm, J., *Mathesis enucleata, or, The elements of the mathematicks*, printed for Robert Knaplock and Dan. Midwinter and Tho. Leigh, London, 1700.

[Tappenden 2005] Tappenden, J., "Proof Style and Understanding in Mathematics I: Visualization, Unification and Axiom Choice", in [Mancosu et al. 2005].

[Teller 1980] Teller, P., "Computer proof", *Journal of Philosophy* 77 (1980), pp. 797–803.

[Tennant 1986] Tennant, N., "The withering away of formal semantics?", *Mind and Language* 1 (1986), pp. 302–318.

[Thomas 2000] Thomas, I., *Selections illustrating the history of Greek mathematics*, vol. II, Aristarchus to Pappus, Loeb Classical Library, Harvard University Press, Cambridge, MA, 2000.

[Tymoczko 1979] Tymoczko, T., "The Four-Color problem and its philosophical significance", *Journal of Philosophy* 76 (1979), pp. 57–83.

[Tymoczko 1980] ——, "Computers, proofs and mathematicians: a philosophical investigation of the four-color proof", *Mathematics Magazine* 53 (1980), pp. 131–138.

[Tymoczko 1986] ——, *New directions in the philosophy of mathematics*, T. Tymoczko (ed.), Birkhäuser, Boston, 1986.

[Viète 1591] Viète, F., *Introduction to the Analytic Art*, ed. and trans. by T. R. Witmer, Kent State University Press, Kent, Ohio, 1983.

[Wallis 1685] Wallis, J., *A Treatise of Algebra, both historical and practical: shewing the original, progress, and advancement thereof, from time to time, and by what steps it hath attained to the height at which it now is.* Printed for Richard Davis by John Playford, London, 1685.

[Wallis 1693] ——, *Opera Mathematica*, Oxoni: E Theatro Sheldoniano, 1693–1699. Reprinted (with a foreword by C. Scriba) by Georg Olms, Hildesheim, New York, 1972.

[Wallis 1656] ——, *The Mathematics of Infinitesimals: John Wallis 1656.* Translated from the Latin by J. Steadall. Springer, New York, 2004.

[Wolff 1739] Wolff, C., *A Treatise of algebra: with the application of it to a variety of problems in arithmetic, to geometry, trigonometry, and conic sections: with the several methods of solving and constructing equations of the higher kind*, printed for A. Bettesworth and C. Hitch, London, 1739.

[Yanney/Calderhead 1898] Yanney, B. & J. Calderhead "New and old proofs of the Pythagorean Theorem", *American Mathematical Monthly* 5 (1898), pp. 73–74.

2

Implications of Experimental Mathematics for the Philosophy of Mathematics[1]

Jonathan Borwein[2]
Faculty of Computer Science
Dalhousie University

From the Editors

When computers were first introduced, they were much more a tool for the other sciences than for mathematics. It was many years before more than a very small subset of mathematicians used them for anything beyond word-processing. Today, however, more and more mathematicians are using computers to actively assist their mathematical research in a range of ways. In this chapter, Jonathan Borwein, one of the leaders in this trend, discusses ways that computers can be used in the development of mathematics, both to assist in the discovery of mathematical facts and to assist in the development of their proofs. He suggests that what mathematics requires is secure *knowledge that mathematical claims are true, and an understanding of why they are true, and that proofs are not necessarily the only route to this security. For teachers of mathematics, computers are a very helpful, if not essential, component of a constructivist approach to the mathematics curriculum.*

Jonathan Borwein holds a Canada Research Chair in the Faculty of Computer Science at Dalhousie University (users.cs.dal.ca/ jborwein/). His research interests include scientific computation, numerical optimization, image reconstruction, computational number theory, experimental mathematics, and collaborative technology. He was the founding Director of the Centre for Experimental and Constructive Mathematics, a Simon Fraser University research center within the Departments of Mathematics and Statistics and Actuarial Science, established in 1993. He has received numerous awards including the Chauvenet Prize of the MAA in 1993 (with

---

[1] The companion web site is at **www.experimentalmath.info**

[2] Canada Research Chair, Faculty of Computer Science, 6050 University Ave, Dalhousie University, Nova Scotia, B3H 1W5 Canada. Email: `jborwein@cs.dal.ca`

*P.B. Borwein and D.H. Bailey) for "Ramanujan, Modular Equations and Pi or How to Compute a Billion Digits of Pi," (*Monthly *1989), Fellowship in the Royal Society of Canada (1994), and Fellowship in the American Association for the Advancement of Science (2002). Jointly with David Bailey he operates the Experimental Mathematics Website, www.experimentalmath.info. He is the author of several hundred papers, and the co-author of numerous books, including, with L. Berggren and P.B. Borwein,* Pi: a Source Book *(Springer-Verlag 1997); with David Bailey,* Mathematics by Experiment: Plausible Reasoning in the 21st Century *(AK Peters 2003); with David Bailey and Roland Girgensohn,* Experiments in Mathematics CD *(AK Peters 2006); with these same co-authors,* Experimentation in Mathematics: Computational Paths to Discovery *(AK Peters 2004); with David Bailey, Neil Calkin, Roland Girgensohn, D. Luke, and Victor Moll,* Experimental Mathematics in Action *(AK Peters 2007[3]); and he has just completed a related book with Keith Devlin,* The Computer as Crucible, *currently in press with AK Peters. Borwein and Bailey have also developed a number of software packages for experimental mathematics (crd.lbl.gov/ dhbailey/expmath/software/).*

Christopher Koch [Koch 2004] accurately captures a great scientific distaste for philosophizing:

> "Whether we scientists are inspired, bored, or infuriated by philosophy, all our theorizing and experimentation depends on particular philosophical background assumptions. This hidden influence is an acute embarrassment to many researchers, and it is therefore not often acknowledged." (Christopher Koch, 2004)

That acknowledged, I am of the opinion that mathematical philosophy matters more now than it has in nearly a century. The power of modern computers matched with that of modern mathematical software and the sophistication of current mathematics is changing the way we do mathematics.

In my view it is now both necessary and possible to admit quasi-empirical inductive methods fully into mathematical argument. In doing so carefully we will enrich mathematics and yet preserve the mathematical literature's deserved reputation for reliability—even as the methods and criteria change. What do I mean by reliability? Well, research mathematicians still consult Euler or Riemann to be informed, anatomists only consult Harvey[4] for historical reasons. Mathematicians happily quote old papers as core steps of arguments, physical scientists expect to have to confirm results with another experiment.

1 Mathematical Knowledge as I View It

Somewhat unusually, I can exactly place the day at registration that I became a mathematician and I recall the reason why. I was about to deposit my punch cards in the 'honours history bin'. I remember thinking

---

[3] An earlier version of this chapter was taught in this short-course based book.

[4] William Harvey published the first accurate description of circulation, "An Anatomical Study of the Motion of the Heart and of the Blood in Animals," in 1628.

> "If I do study history, in ten years I shall have forgotten how to use the calculus properly. If I take mathematics, I shall still be able to read competently about the War of 1812 or the Papal schism." (Jonathan Borwein, 1968)

The inescapable reality of objective mathematical knowledge is still with me. Nonetheless, my view then of the edifice I was entering is not that close to my view of the one I inhabit forty years later.

I also know when I became a computer-assisted fallibilist. Reading Imre Lakatos' *Proofs and Refutations*, [Lakatos 1976], a few years later while a very new faculty member, I was suddenly absolved from the grave sin of error, as I began to understand that missteps, mistakes and errors are the grist of all creative work.[5] The book, his doctorate posthumously published in 1976, is a student conversation about the Euler characteristic. The students are of various philosophical stripes and the discourse benefits from his early work on Hegel with the Stalinist Lukács in Hungary and from later study with Karl Popper at the London School of Economics. I had been prepared for this dispensation by the opportunity to learn a variety of subjects from Michael Dummett. Dummett was at that time completing his study rehabilitating Frege's status, [Dummett 1973].

A decade later the appearance of the first 'portable' computers happily coincided with my desire to decode Srinivasa Ramanujan's (1887–1920) cryptic assertions about theta functions and elliptic integrals, [Borwein et al. 1989]. I realized that by coding his formulae and my own in the *APL* programming language[6], I was able to rapidly confirm and refute identities and conjectures and to travel much more rapidly and fearlessly down potential blind alleys. I had become a computer-assisted fallibilist, at first somewhat falteringly, but twenty years have certainly honed my abilities.

Today, while I appreciate fine proofs and aim to produce them when possible, I no longer view proof as the royal road to secure mathematical knowledge.

2 Introduction

I first discuss my views, and those of others, on the nature of mathematics, and then illustrate these views in a variety of mathematical contexts. A considerably more detailed treatment of many of these topics is to be found in my book with Dave Bailey entitled *Mathematics by Experiment: Plausible Reasoning in the 21st Century*—especially in Chapters One, Two and Seven, [Borwein/Bailey 2003]. Additionally, [Bailey et al. 2007] contains several pertinent case studies as well as a version of this current chapter.

Kurt Gödel may well have overturned the mathematical apple cart entirely deductively, but nonetheless he could hold quite different ideas about legitimate forms of mathematical reasoning, [Gödel 1995]:

> "If mathematics describes an objective world just like physics, there is no reason why inductive methods should not be applied in mathematics just the same as in physics." (Kurt Gödel[7], 1951)

[5] Gila Hanna [Hanna 2006] takes a more critical view placing more emphasis on the role of proof and certainty in mathematics; I do not disagree, so much as I place more value on the role of computer-assisted refutation. Also 'certainty' usually arrives late in the development of a proof.

[6] Known as a 'write only' very high level language, APL was a fine tool, albeit with a steep learning curve whose code is almost impossible to read later.

[7] Taken from a previously unpublished work, [Gödel 1995] originally given as the 1951 Gibbs lecture.

While we mathematicians have often separated ourselves from the sciences, they have tended to be more ecumenical. For example, a recent review of *Models. The Third Dimension of Science*, [Brown 2004], chose a mathematical plaster model of a Clebsch diagonal surface as its only illustration. Similarly, authors seeking examples of the aesthetic in science often choose iconic mathematics formulae such as $E = MC^2$.

Let me begin by fixing a few concepts before starting work in earnest. Above all, I hope to persuade you of the power of mathematical experimentation—it is also fun—and that the traditional accounting of mathematical learning and research is largely an ahistorical caricature. I recall three terms.

mathematics, n. *a group of related subjects, including algebra, geometry, trigonometry and calculus, concerned with the study of number, quantity, shape, and space, and their inter-relationships, applications, generalizations and abstractions.*

This definition—taken from my Collins Dictionary [Borowski/Borwein 2006]—makes no immediate mention of proof, nor of the means of reasoning to be allowed. The Webster's Dictionary [Webster's 1999] contrasts:

induction, n. *any form of reasoning in which the conclusion, though supported by the premises, does not follow from them necessarily.*; and

deduction, n. *a process of reasoning in which a conclusion follows necessarily from the premises presented, so that the conclusion cannot be false if the premises are true.*
b. a conclusion reached by this process.

Like Gödel, I suggest that both should be entertained in mathematics. This is certainly compatible with the general view of mathematicians that in some sense "mathematical stuff is out there" to be discovered. In this paper, I shall talk broadly about experimental and heuristic mathematics, giving accessible, primarily visual and symbolic, examples.

3 Philosophy of Experimental Mathematics

> "The computer has in turn changed the very nature of mathematical experience, suggesting for the first time that mathematics, like physics, may yet become an empirical discipline, a place where things are discovered because they are seen."
>
> (David Berlinski, [Berlinski 1997], p. 39)

The shift from *typographic* to *digital culture* is vexing for mathematicians. For example, there is still no truly satisfactory way of displaying mathematics on the web—and certainly not of asking mathematical questions. Also, we respect *authority*, [Grabiner 2004], but value *authorship* deeply—however much the two values are in conflict, [Borwein/Stanway 2005]. For example, the more I recast someone else's ideas in my own words, the more I enhance my authorship while undermining the original authority of the notions. Medieval scribes had the opposite concern and so took care to attribute their ideas to such as Aristotle or Plato.

And we care more about the *reliability* of our literature than does any other science. Indeed I would argue that we have over-subscribed to this notion and often pay lip-service, not real attention, to our older literature. How often does one see original sources sprinkled like holy water in papers that make no real use of them—the references offering a false sense of scholarship?

The traditional central role of proof in mathematics is arguably and perhaps appropriately under siege. Via examples, I intend to pose and answer various questions. I shall conclude with a variety of quotations from our progenitors and even contemporaries:

My Questions. What constitutes secure mathematical knowledge? When is computation convincing? Are humans less fallible? What tools are available? What methodologies? What of the 'law of the small numbers'? Who cares for certainty? What is the role of proof? How is mathematics actually done? How should it be? I mean these questions both about the apprehension (discovery) and the establishment (proving) of mathematics. This is presumably more controversial in the formal proof phase.

My Answers. To misquote D'Arcy Thompson (1860–1948) 'form follows function', [Thompson 1992]: rigour (proof) follows reason (discovery); indeed, excessive focus on rigour has driven us away from our wellsprings. Many good ideas are wrong. Not all truths are provable, and not all provable truths are worth proving. Gödel's incompleteness results certainly showed us the first two of these assertions while the third is the bane of editors who are frequently presented with correct but unexceptional and unmotivated generalizations of results in the literature. Moreover, near certainty is often as good as it gets—intellectual context (community) matters. Recent complex human proofs are often very long, extraordinarily subtle and fraught with error—consider Fermat's last theorem, the Poincaré conjecture, the classification of finite simple groups, presumably any proof of the Riemann hypothesis, [Economist 2005]. So while we mathematicians publicly talk of certainty we really settle for security.

In all these settings, modern computational tools dramatically change the nature and scale of available evidence. Given an interesting identity buried in a long and complicated paper on an unfamiliar subject, which would give you more confidence in its correctness: staring at the proof, or confirming computationally that it is correct to 10,000 decimal places?

Here is such a formula ([Bailey/Borwein 2005], p. 20):

$$\frac{24}{7\sqrt{7}}\int_{\pi/3}^{\pi/2}\log\left|\frac{\tan t+\sqrt{7}}{\tan t-\sqrt{7}}\right|dt \stackrel{?}{=} L_{-7}(2)$$

$$=\sum_{n=0}^{\infty}\left[\frac{1}{(7n+1)^2}+\frac{1}{(7n+2)^2}-\frac{1}{(7n+3)^2}+\frac{1}{(7n+4)^2}-\frac{1}{(7n+5)^2}-\frac{1}{(7n+6)^2}\right]. \quad (1)$$

This identity links a volume (the integral) to an arithmetic quantity (the sum). It arose out of some studies in quantum field theory, in analysis of the volumes of ideal tetrahedra in hyperbolic space. The question mark is used because, while no hint of a path to a formal proof is yet known, it has been verified numerically to 20,000 digit precision—using 45 minutes on 1024 processors at Virginia Tech.

A more inductive approach can have significant benefits. For example, as there is still some doubt about the proof of the classification of finite simple groups it is important to ask whether the result is true but the proof flawed, or rather if there is still perhaps an 'ogre' sporadic group even larger than the 'monster.' What heuristic, probabilistic or computational tools can increase our confidence that the ogre does or does not exist? Likewise, there are experts who still believe the *Riemann hypothesis*[8] (RH) may be false and that the billions of zeroes found so far are much too small to be representative.[9] In any event, our understanding of the complexity of various crypto-systems relies on (RH) and we should like secure knowledge that any counter-example is enormous.

Peter Medawar (1915–87)—a Nobel prize winning oncologist and a great expositor of science—writing in *Advice to a Young Scientist*, [Medawar 1979], identifies four forms of scientific experiment:

1. *The Kantian experiment: generating "the classical non-Euclidean geometries (hyperbolic, elliptic) by replacing Euclid's axiom of parallels (or something equivalent to it) with alternative forms."* All mathematicians perform such experiments while the majority of computer explorations are of the following Baconian form.
2. *The Baconian experiment is a contrived as opposed to a natural happening, it "is the consequence of 'trying things out' or even of merely messing about."* Baconian experiments are the explorations of a happy if disorganized beachcomber and carry little predictive power.
3. *Aristotelian demonstrations: "apply electrodes to a frog's sciatic nerve, and lo, the leg kicks; always precede the presentation of the dog's dinner with the ringing of a bell, and lo, the bell alone will soon make the dog dribble."* Arguably our 'Corollaries' and 'Examples' are Aristotelian, they reinforce but do not predict. Medawar then says the most important form of experiment is:
4. *The Galilean experiment is "a critical experiment—one that discriminates between possibilities and, in doing so, either gives us confidence in the view we are taking or makes us think it in need of correction."* The Galilean is the only form of experiment which stands to make Experimental Mathematics a serious enterprise. Performing careful, replicable Galilean experiments requires work and care.

Reuben Hersh's arguments for a humanist philosophy of mathematics, especially ([Hersh 1995], pp. 590–591), and ([Hersh 1999], p. 22), as paraphrased below, become even more convincing in our highly computational setting.

1. Mathematics is human. *It is part of and fits into human culture. It does not match Frege's concept of an abstract, timeless, tenseless, objective reality.*[10]
2. Mathematical knowledge is fallible. *As in science, mathematics can advance by making mistakes and then correcting or even re-correcting them. The "fallibilism" of mathematics is brilliantly argued in Lakatos' Proofs and Refutations.*
3. There are different versions of proof or rigor. *Standards of rigor can vary depending on time, place, and other things. The use of computers in formal proofs, exemplified by the*

---

[8] All non-trivial zeroes—not negative even integers—of the zeta function lie on the line with real part 1/2.

[9] See [Odlyzko 2001] and various of Andrew Odlyzko's unpublished but widely circulated works.

[10] That Frege's view of mathematics is wrong, for Hersh as for me, does not diminish its historical importance.

computer-assisted proof of the four color theorem in 1977,[11] *is just one example of an emerging nontraditional standard of rigor.*

4. Empirical evidence, numerical experimentation and probabilistic proof all can help us decide what to believe in mathematics. *Aristotelian logic isn't necessarily always the best way of deciding.*
5. Mathematical objects are a special variety of a social-cultural-historical object. *Contrary to the assertions of certain post-modern detractors, mathematics cannot be dismissed as merely a new form of literature or religion. Nevertheless, many mathematical objects can be seen as shared ideas, like Moby Dick in literature, or the Immaculate Conception in religion.*

I entirely subscribe to points 2., 3., 4., and with certain caveats about objective knowledge[12] to points 1. and 5. In any event mathematics is and will remain a uniquely human undertaking.

This version of humanism sits *fairly* comfortably along-side current versions of **social-constructivism** as described next.

> "The social constructivist thesis is that mathematics is a social construction, a cultural product, fallible like any other branch of knowledge." (Paul Ernest, [Ernest 1990], §3)

But only if I qualify this with "*Yes, but much-much less fallible than most branches of knowledge.*" Associated most notably with the writings of Paul Ernest—an English Mathematician and Professor in the Philosophy of Mathematics Education who in [Ernest 1998] traces the intellectual pedigree for his thesis, a pedigree that encompasses the writings of Wittgenstein, Lakatos, Davis, and Hersh among others—social constructivism seeks to define mathematical knowledge and epistemology through the social structure and interactions of the mathematical community and society as a whole.

This interaction often takes place over very long periods. Many of the ideas our students—and some colleagues—take for granted took a great deal of time to gel. The Greeks suspected the impossibility of the three *classical construction problems*[13] and the irrationality of the golden mean was well known to the Pythagoreans.

While concerns about potential and completed infinities are very old, until the advent of the calculus with Newton and Leibnitz and the need to handle fluxions or infinitesimals, the level of need for rigour remained modest. Certainly Euclid is in its geometric domain generally a model of rigour, while also Archimedes' numerical analysis was not equalled until the 19th century.

The need for rigour arrived in full force in the time of Cauchy and Fourier. The treacherous countably infinite processes of analysis and the limitations of formal manipulation came to the fore. It is difficult with a modern sensibility to understand how Cauchy's proof of the continuity

---

[11] Especially since a new implementation by Seymour, Robertson and Thomas in 1997 has produced a simpler, clearer and less troubling implementation.

[12] While it is not Hersh's intention, a superficial reading of point 5. hints at a cultural relativism to which I certainly do not subscribe.

[13] Trisection, circle squaring and cube doubling were taken by the educated to be impossible in antiquity. Already in 414 BCE, in his play *The Birds*, Aristophanes uses 'circle-squarers' as a term for those who attempt the impossible. Similarly, the French Academy stopped accepting claimed proofs a full two centuries before the 19th century achieved proofs of their impossibility.

of pointwise-limits could coexist in texts for a generation with clear counter-examples originating in Fourier's theory of heat.[14]

By the end of the 19th century Frege's (1848–1925) attempt to base mathematics in a linguistically based *logicism* had foundered on Russell and other's discoveries of the paradoxes of naive set theory. Within thirty five years Gödel—and then Turing's more algorithmic treatment[15]—had similarly damaged both Russell and Whitehead's and Hilbert's programs.

Throughout the twentieth century, bolstered by the armor of abstraction, the great ship Mathematics has sailed on largely unperturbed. During the last decade of the 19th and first few decades of the 20th century the following main streams of philosophy emerged explicitly within mathematics to replace logicism, but primarily as the domain of philosophers and logicians.

- *Platonism.* Everyman's idealist philosophy—stuff exists and we must find it. Despite being the oldest mathematical philosophy, Platonism—still predominant among working mathematicians—was only christened in 1934 by Paul Bernays.[16]
- *Formalism.* Associated mostly with Hilbert—it asserts that mathematics is invented and is best viewed as formal symbolic games without intrinsic meaning.
- *Intuitionism.* Invented by Brouwer and championed by Heyting, intuitionism asks for inarguable monadic components that can be fully analyzed and has many variants; this has interesting overlaps with recent work in cognitive psychology such as Lakoff and Nunez' work, [Lakoff/Nunez 2001], on 'embodied cognition'.[17]
- *Constructivism.* Originating with Markoff and especially Kronecker (1823–1891), and refined by Bishop it finds fault with significant parts of classical mathematics. Its 'I'm from Missouri, tell me how big it is' sensibility is not to be confused with Paul Ernest's 'social constructivism', [Ernest 1998].

The last two philosophies deny the principle of the *excluded middle*, "*A* or not *A*," and resonate with computer science—as does some of formalism. It is hard after all to run a deterministic program which does not know which disjunctive logic-gate to follow. By contrast the battle between a Platonic idealism (a 'deductive absolutism') and various forms of 'fallibilism' (a quasi-empirical 'relativism') plays out across all four, but fallibilism perhaps lives most easily within a restrained version of intuitionism which looks for 'intuitive arguments' and is willing to accept that 'a proof is what convinces'. As Lakatos shows, an argument that was convincing a hundred years ago may well now be viewed as inadequate. And one today trusted may be challenged in the next century.

---

[14] Cauchy's proof appeared in his 1821 text on analysis. While counterexamples were pointed out almost immediately, Stokes and Seidel were still refining the missing uniformity conditions in the late 1840s.

[15] The modern treatment of incompleteness leans heavily on Turing's analysis of the *Halting problem* for so-called Turing machines.

[16] See Karlis Podnieks, "Platonism, Intuition and the Nature on Mathematics," available at `www.ltn.lv/podnieks/gt1.html`

[17] The cognate views of Henri Poincaré (1854–1912) ([Poincaré 2004], p. 23) on the role of the *subliminal* are reflected in "The mathematical facts that are worthy of study are those that, by their analogy with other facts are susceptible of leading us to knowledge of a mathematical law, in the same way that physical facts lead us to a physical law." He also wrote "It is by logic we prove, it is by intuition that we invent," [Poincaré 1904].

As we illustrate in the next section or two, it is only perhaps in the last twenty five years, with the emergence of powerful mathematical platforms, that any approach other than a largely undigested Platonism and a reliance on proof and abstraction has had the tools[18] to give it traction with working mathematicians.

In this light, Hales' proof of Kepler's conjecture that *the densest way to stack spheres is in a pyramid* resolves the oldest problem in discrete geometry. It also supplies the most interesting recent example of intensively computer-assisted proof, and after five years with the review process was published in the *Annals of Mathematics*—with an "only 99% checked" disclaimer, withdrawn very late in the process and after being widely reported.

This process has triggered very varied reactions [Kolata 2004] and has provoked Thomas Hales to attempt a formal computational proof which he expects to complete by 2011, [Economist 2005]. Famous earlier examples of fundamentally computer-assisted proof include the *Four color theorem* and proof of the *Non-existence of a projective plane of order 10*. The three raise and answer quite distinct questions about computer-assisted proof—both real and specious. For example, there were real concerns about the completeness of the search in the 1976 proof of the Four color theorem but there should be none about the 1997 reworking by Seymour, Robertson and Thomas.[19] Correspondingly, Lam deservedly won the 1992 *Lester R. Ford award* for his compelling explanation of why to trust his computer when it announced there was no plane of order ten, [Lam 1991]. Finally, while it is reasonable to be concerned about the certainty of Hales' conclusion, was it really the *Annal's* purpose to suggest all other articles have been more than 99% certified?

To make the case as to how far mathematical computation has come we trace the changes over the past half century. The 1949 computation of π to 2,037 places suggested by von Neumann, took 70 hours. A billion digits may now be computed in much less time on a laptop. Strikingly, it would have taken roughly 100,000 ENIAC's to store the Smithsonian's picture—as is possible thanks to *40 years of Moore's law* in action.[20]

This is an astounding record of sustained exponential progress without peer in the history of technology. Additionally, mathematical tools are now being implemented on parallel platforms, providing *much* greater power to the research mathematician. Amassing huge amounts of processing power will not alone solve many mathematical problems. There are very few mathematical 'Grand-challenge problems', [JBorwein/PBorwein 2001] where, as in the physical sciences, a few more orders of computational power will resolve a problem.

For example, an order of magnitude improvement in computational power currently translates into one more day of accurate weather forecasting, while it is now common for biomedical researchers to design experiments today whose outcome is predicated on 'peta-scale' computation being available by say 2010, [Rowe et al. 2005]. There is, however, much more value in *very rapid 'Aha's'* as can be obtained through "micro-parallelism;" that is, where we benefit by being able to compute many simultaneous answers on a neurologically-rapid scale and so can hold many parts of a problem in our mind at one time.

---

[18] That is, to broadly implement Hersh's central points (2.-4.).

[19] See `www.math.gatech.edu/thomas/FC/fourcolor.html`.

[20] **Moore's Law** is now taken to be the assertion that *semiconductor technology approximately doubles in capacity and performance roughly every 18 to 24 months.*

To sum up, in light of the discussion and terms above, I now describe myself a sort-of social-constructivist, and as a computer-assisted fallibilist with constructivist leanings. I believe that more-and-more of the interesting parts of mathematics will be less-and-less susceptible to classical deductive analysis and that Hersh's 'non-traditional standard of rigor' must come to the fore.

4 Our Experimental Mathodology

Despite Picasso's complaint that "computers are useless, they only give answers," the main goal of computation in pure mathematics is arguably to yield *insight*. This demands speed or, equivalently, substantial *micro-parallelism* to provide answers on a cognitively relevant scale; so that we may ask and answer more questions while they remain in our consciousness. This is relevant for rapid verification; for validation; for *proofs* and *especially for refutations* which includes what Lakatos calls "monster barring," [Lakatos 1976]. Most of this goes on in the daily small-scale accretive level of mathematical discovery but insight is gained even in cases like the proof of the Four color theorem or the Non-existence of a plane of order ten. Such insight is not found in the case-enumeration of the proof, but rather in the algorithmic reasons for believing that one has at hand a tractable unavoidable set of configurations or another effective algorithmic strategy. For instance, Lam [Lam 1991] ran his algorithms on known cases in various subtle ways, and also explained why built-in redundancy made the probability of machine-generated error negligible. More generally, the act of programming—if well performed—always leads to more insight about the structure of the problem.

In this setting it is enough to equate *parallelism* with access to requisite *more* space and speed of computation. Also, we should be willing to consider all computations as 'exact' which provide truly reliable answers.[21] This now usually requires a careful *hybrid* of symbolic and numeric methods, such as achieved by *Maple*'s liaison with the *Numerical Algorithms Group* (NAG) Library[22], see [Bornemann et al. 2004], [Borwein 2005b]. There are now excellent tools for such purposes throughout analysis, algebra, geometry and topology, see [Borwein/Bailey 2003], [Borwein et al. 2004], [Bornemann et al. 2004], [JBorwein/PBorwein 2001], [Borwein/Corless 1999].

Along the way questions required by—or just made natural by—computing start to force out older questions and possibilities in the way beautifully described a century ago by Dewey regarding evolution.

> "Old ideas give way slowly; for they are more than abstract logical forms and categories. They are habits, predispositions, deeply engrained attitudes of aversion and preference. Moreover, the conviction persists—though history shows it to be a hallucination—that all the questions that the human mind has asked are questions that can be answered in terms of the alternatives that the questions themselves present. But in fact intellectual progress usually occurs through sheer abandonment of questions together with both of the alternatives they assume; an abandonment that results from their decreasing vitality and a change of urgent interest. We do not solve them: we get over them. Old questions

---

[21] If careful interval analysis can certify that a number known to be integer is larger than 2.5 and less than 3.5, this constitutes an exact computational proof that it is 3.

[22] See `www.nag.co.uk/`.

are solved by disappearing, evaporating, while new questions corresponding to the changed attitude of endeavor and preference take their place. Doubtless the greatest dissolvent in contemporary thought of old questions, the greatest precipitant of new methods, new intentions, new problems, is the one effected by the scientific revolution that found its climax in the 'Origin of Species.'" (John Dewey, [Dewey 1997])

Lest one think this a feature of the humanities and the human sciences, consider the artisanal chemical processes that have been lost as they were replaced by cheaper industrial versions. And mathematics is far from immune. Felix Klein, quoted at length in the introduction to [JBorwein/PBorwein 1987], laments that "now the younger generation hardly knows abelian functions." He goes on to explain that:

"In mathematics as in the other sciences, the same processes can be observed again and again. First, new questions arise, for internal or external reasons, and draw researchers away from the old questions. And the old questions, just because they have been worked on so much, need ever more comprehensive study for their mastery. This is unpleasant, and so one is glad to turn to problems that have been less developed and therefore require less foreknowledge—even if it is only a matter of axiomatics, or set theory, or some such thing." (Felix Klein, [Klein 1928], p. 294)

Freeman Dyson has likewise gracefully described how taste changes:

"I see some parallels between the shifts of fashion in mathematics and in music. In music, the popular new styles of jazz and rock became fashionable a little earlier than the new mathematical styles of chaos and complexity theory. Jazz and rock were long despised by classical musicians, but have emerged as art-forms more accessible than classical music to a wide section of the public. Jazz and rock are no longer to be despised as passing fads. Neither are chaos and complexity theory. But still, classical music and classical mathematics are not dead. Mozart lives, and so does Euler. When the wheel of fashion turns once more, quantum mechanics and hard analysis will once again be in style." (Freeman Dyson, [Dyson 1996])

For example recursively defined objects were once anathema—Ramanujan worked very hard to replace lovely iterations by sometimes-obscure closed-form approximations. Additionally, what is "easy" changes: high performance computing and networking are blurring, merging disciplines and collaborators. This is democratizing mathematics but further challenging authentication—consider how easy it is to find information on *Wikipedia*[23] and how hard it is to validate it.

Moving towards a well articulated Experimental *Mathodology*—both in theory and practice—will take much effort. The need is premised on the assertions that intuition is acquired—we can and must better mesh computation and mathematics, and that visualization is of growing importance—in many settings even three is a lot of dimensions.

[23] *Wikipedia* is an open source project at `en.wikipedia.org/wiki/Main_Page`; "wiki-wiki" is Hawaiian for "quickly."

"Monster-barring" (Lakatos's term, [Lakatos 1976], for refining hypotheses to rule out nasty counter-examples[24]) and "caging" (Nathalie Sinclair tells me this is my own term for imposing needed restrictions in a conjecture) are often easy to enhance computationally, as for example with randomized checks of equations, linear algebra, and primality or graphic checks of equalities, inequalities, areas, etc. Moreover, our mathodology fits well with the kind of pedagogy espoused at a more elementary level (and without the computer) by John Mason in [Mason 2006].

4.1 Eight Roles for Computation

I next recapitulate eight roles for computation that Bailey and I discuss in our two recent books [Borwein/Bailey 2003], [Borwein et al. 2004]:

#1. Gaining insight and intuition or just knowledge. Working algorithmically with mathematical objects almost inevitably adds insight to the processes one is studying. At some point even just the careful aggregation of data leads to better understanding.

#2. Discovering new facts, patterns and relationships. The number of *additive partitions* of a positive integer n, $p(n)$, is *generated* by

$$P(q) := 1 + \sum_{n \geq 1} p(n) q^n = \frac{1}{\prod_{n=1}^{\infty}(1 - q^n)}. \tag{2}$$

Thus, $p(5) = 7$ since

$$5 = 4 + 1 = 3 + 2 = 3 + 1 + 1 = 2 + 2 + 1 = 2 + 1 + 1 + 1 = 1 + 1 + 1 + 1 + 1.$$

Developing (2) is a fine introduction to enumeration via *generating functions*. Additive partitions are harder to handle than multiplicative factorizations, but they are very interesting ([Borwein et al. 2004], Chapter 4). Ramanujan used Major MacMahon's table of $p(n)$ to intuit remarkable deep congruences such as

$$p(\mathbf{5n+4}) \equiv 0 \mod \mathbf{5}, \quad p(\mathbf{7n+5}) \equiv 0 \mod \mathbf{7}, \quad p(\mathbf{11n+6}) \equiv 0 \mod \mathbf{11},$$

from relatively limited data like

$$\begin{aligned} P(q) = {} & 1 + q + 2q^2 + 3q^3 + \underline{5}q^{\mathbf{4}} + \overline{7}q^{\mathbf{5}} + 11q^6 + 15q^7 \\ & + 22q^8 + \underline{30}q^{\mathbf{9}} + 42q^{10} + 56q^{11} + \overline{77}q^{\mathbf{12}} + 101q^{13} + \underline{135}q^{\mathbf{14}} \\ & + 176q^{15} + 231q^{16} + 297q^{17} + 385q^{18} + \underline{\overline{490}}q^{\mathbf{19}} \\ & + 627q^{20}b + 792q^{21} + 1002q^{22} + \cdots + p(200)q^{200} + \cdots \end{aligned} \tag{3}$$

Cases $5n + 4$ and $7n + 5$ are flagged in (3). Of course, it is markedly easier to (heuristically) confirm than find these fine examples of *Mathematics: the science of patterns.*[25] The study of such congruences—much assisted by symbolic computation—is very active today.

---

[24] Is, for example, a polyhedron always convex? Is a curve intended to be simple? Is a topology assumed Hausdorff, a group commutative?

[25] The title of Keith Devlin's 1996 book, [Devlin 1996].

#3. Graphing to expose mathematical facts, structures or principles. Consider Nick Trefethen's fourth challenge problem as described in [Bornemann et al. 2004], [Borwein 2005b]. It requires one to find ten good digits of:

4. What is the global minimum of the function

$$\begin{aligned}&\exp(\sin(50x)) + \sin(60e^{y}) + \sin(70\sin x) + \sin(\sin(80y))\\&\quad - \sin(10(x+y)) + (x^2+y^2)/4?\end{aligned}$$

As a foretaste of future graphic tools, one can solve this problem graphically and interactively using current *adaptive 3-D plotting* routines which can catch all the bumps. This does admittedly rely on trusting a good deal of software.

#4. Rigourously testing and especially falsifying conjectures. I hew to the Popperian scientific view that we primarily falsify; but that as we perform more and more testing experiments without such falsification we draw closer to firm belief in the truth of a conjecture such as: *the polynomial* $P(n) = n^2 - n + p$ *has prime values for all* $n = 0, 1, \ldots, p-2$, *exactly for* Euler's lucky prime numbers, *that is,* $p = 2, 3, 5, 11, 17,$ *and* 41.[26]

#5. Exploring a possible result to see if it *merits* formal proof. A conventional deductive approach to a hard multi-step problem really requires establishing all the subordinate lemmas and propositions needed along the way—especially if they are highly technical and un-intuitive. Now some may be independently interesting or useful, but many are only worth proving if the entire expedition pans out. Computational experimental mathematics provides tools to survey the landscape with little risk of error: only if the view from the summit is worthwhile, does one lay out the route carefully. I discuss this further at the end of the next Section.

#6. Suggesting approaches for formal proof. The proof of the *cubic theta function identity* discussed in ([Borwein et al. 2004], p. 210ff), shows how a fully intelligible human proof can be obtained entirely by careful symbolic computation.

#7. Computing replacing lengthy hand derivations. Who would wish to verify the following prime factorization by hand?

$$\begin{aligned}&6422607578676942838792549775208734746307\\&= (2140992015395526641)(1963506722254397)(1527791).\end{aligned}$$

Surely, what we value is understanding the underlying algorithm, not the human work?

#8. Confirming analytically derived results. This is a wonderful and frequently accessible way of confirming results. Even if the result itself is not computationally checkable, there is often an accessible corollary. An assertion about bounded operators on Hilbert space may have a useful consequence for three-by-three matrices. It is also an excellent way to error correct, or to check calculus examples before giving a class.

5 Finding Things versus Proving Things

I now illuminate these eight roles with eight mathematical examples. At the end of each I note some of the roles illustrated.

[26] See [Weisstein WWW] for the answer.

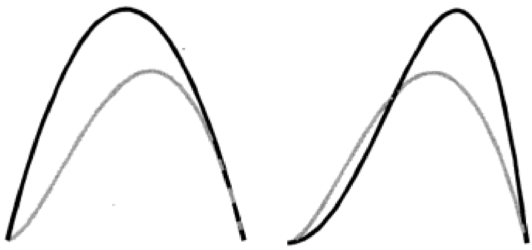


Figure 2.1 **(Ex. 1.): Graphical comparison of** $-x^2 \ln(x)$ (lower local maximum in both graphs) with $x - x^2$ (left graph) and $x^2 - x^4$ (right graph)

1. **Pictorial comparison** of $y - y^2$ and $y^2 - y^4$ to $-y^2 \ln(y)$, when y lies in the unit interval, is a much more rapid way to divine which function is larger than by using traditional analytic methods.

   Figure 2.1 below shows that it is clear in the latter case that the functions cross, and so it is futile to try to prove one majorizes the other. In the first case, evidence is provided to motivate attempting a proof and often the picture serves to guide such a proof—by showing monotonicity or convexity or some other salient property. ■

This certainly illustrates roles #3 and #4, and perhaps role #5.

2. **A proof and a disproof.** Any modern computer algebra can tell one that

$$0 < \int_0^1 \frac{(1-x)^4 x^4}{1+x^2}\,dx = \frac{22}{7} - \pi, \tag{4}$$

   since the integral may be interpreted as the area under a positive curve. We are however no wiser as to why! If however we ask the same system to compute the indefinite integral, we are likely to be told that

$$\int_0^t \cdot = \frac{1}{7}t^7 - \frac{2}{3}t^6 + t^5 - \frac{4}{3}t^3 + 4\,t - 4\,\arctan(t)\,.$$

   Then (4) is now rigourously established by differentiation and an appeal to the Fundamental theorem of calculus. ■

This illustrates roles #1 and #6. It also falsifies the bad conjecture that $\pi = 22/7$ and so illustrates #4 again. Finally, the computer's proof is easier (#7) and very nice, though probably it is not the one we would have developed by ourselves. The fact that $22/7$ is a continued fraction approximation to π has led to many hunts for generalizations of (4), see [Borwein et al. 2004], Chapter 1. None so far are entirely successful.

3. **A computer discovery and a 'proof' of the series for** $\arcsin^2(x)$. We compute a few coefficients and observe that there is a regular power of 4 in the numerator, and integers in the denominator; or equivalently we look at $\arcsin(x/2)^2$. The generating function package 'gfun' in *Maple*, then predicts a recursion, r, for the denominators and solves it, as R.

```
>with(gfun):
>s:=[seq(1/coeff(series(arcsin(x/2)^2,x,25),x,2*n),n=1..6)]:
>R:=unapply(rsolve(op(1, listtorec(s,r(m))),r(m)),m);[seq(R(m),m=0..8)];
```

yields, $s := [4, 48, 360, 2240, 12600, 66528]$,

$$R := m \mapsto 8\,\frac{4^m\,\Gamma(3/2+m)(m+1)}{\pi^{1/2}\Gamma(1+m)},$$

where Γ is the Gamma function, and then returns the sequence of values

$$[4, 48, 360, 2240, 12600, 66528, 336336, 1647360, 7876440].$$

We may now use Sloane's *Online Encyclopedia of Integer Sequences*[27] to reveal that the coefficients are $R(n) = 2n^2\binom{2n}{n}$. More precisely, sequence A002544 identifies $R(n+1)/4 = \binom{2n+1}{n}(n+1)^2$.

```
> [seq(2*n^2*binomial(2*n,n),n=1..8)];
```

confirms this with

$$[4, 48, 360, 2240, 12600, 66528, 336336, 1647360].$$

Next we write

```
> S:=Sum((2*x)^(2*n)/(2*n^2*binomial(2*n,n)),n=1..infinity):S=values(S);
```

which returns

$$\frac{1}{2}\sum_{n=1}^{\infty}\frac{(2\,x)^{2n}}{n^2\binom{2n}{n}} = \arcsin^2(x).$$

That is, we have discovered—and proven if we trust or verify *Maple*'s summation algorithm—the desired Maclaurin series.

As prefigured by Ramanujan, it transpires that there is a beautiful closed form for $\arcsin^{2m}(x)$ for all $m = 1, 2, \ldots$. In [Borwein/Chamberland 2007] there is a discussion of the use of *integer relation methods*, [Borwein/Bailey 2003], Chapter 6, to find this closed form and associated proofs are presented. ■

Here we see an admixture of all of the roles save #3, but above all #2 and #5.

4. **Discovery without proof.** Donald Knuth[28] asked for a closed form evaluation of:

$$\sum_{k=1}^{\infty}\left\{\frac{k^k}{k!\,e^k} - \frac{1}{\sqrt{2\,\pi\,k}}\right\} = -0.084069508727655\ldots. \tag{5}$$

Since about 2000 CE it has been easy to compute 20—or 200—digits of this sum in *Maple* or *Mathematica*; and then to use the 'smart lookup' facility in the *Inverse Symbolic Calculator*(ISC). The ISC at `oldweb.cecm.sfu.ca/projects/ISC` uses a variety of search algorithms and heuristics to predict what a number might actually be. Similar ideas are now implemented as 'identify' in *Maple* and (for algebraic numbers only) as 'Recognize' in *Mathematica*, and are described in [Borwein 2005b], [Borwein/Bailey 2003],

[27] At `www.research.att.com/~njas/sequences/index.html`.

[28] Posed as an MAA Problem [Knuth 2002].

[Borwein/Corless 1999], [Bailey/Borwein 2000]. In this case it *rapidly* returns

$$0.084069508727655 \approx \frac{2}{3} + \frac{\zeta(1/2)}{\sqrt{2\pi}}.$$

We thus have a prediction which *Maple* 9.5 on a 2004 laptop *confirms* to 100 places in under 6 seconds and to 500 in 40 seconds. Arguably we are done. After all we were asked to *evaluate* the series and we now know a closed-form answer.

Notice also that the 'divergent' $\zeta(1/2)$ term is formally to be expected in that while $\sum_{n=1}^{\infty} 1/n^{1/2} = \infty$, the *analytic continuation* of $\zeta(s) := \sum_{n=1}^{\infty} 1/n^s$ for $s > 1$ evaluated at $1/2$ does occur! ■

We have discovered and tested the result and in so doing gained insight and knowledge while illustrating roles #1, #2 and #4. Moreover, as described in [Borwein et al. 2004], p. 15, one can also be led by the computer to a very satisfactory computer-assisted but also very human proof, thus illustrating role #6. Indeed, the first hint is that the computer algebra system returned the value in (5) very quickly even though the series is very slowly convergent. This suggests the program is doing something intelligent—and it is! Such a use of computing is termed "instrumental" in that the computer is fundamental to the process, see [Lagrange 2005].

5. **A striking conjecture with no known proof strategy** (as of spring 2007)[29] given in [Borwein et al. 2004], p. 162, is: for $n = 1, 2, 3 \ldots$

$$8^n \zeta\left(\{\overline{2}, 1\}_n\right) \stackrel{?}{=} \zeta\left(\{2, 1\}_n\right). \tag{6}$$

Explicitly, the first two cases are

$$8 \sum_{n>m>0} \frac{(-1)^n}{n^2 m} = \sum_{n>0} \frac{1}{n^3} \quad \text{and} \quad 64 \sum_{n>m>o>p>0} \frac{(-1)^{n+o}}{n^2 m\, o^2 p} = \sum_{n>m>0} \frac{1}{n^3 m^3}.$$

The notation should now be clear—we use the 'overbar' to denote an alternation. Such alternating sums are called *multi-zeta values* (MZV) and positive ones are called *Euler sums* after Euler who first studied them seriously. They arise naturally in a variety of modern fields from combinatorics to mathematical physics and knot theory.

There is abundant evidence amassed since 'identity' (6) was found in 1996. For example, very recently Petr Lisonek checked the first 85 cases to 1000 places in about 41 HP hours with only the *predicted round-off error*. And the case $n = 163$ was checked in about ten hours. These objects are very hard to compute naively and require substantial computation as a precursor to their analysis.

Formula (6) is the *only* identification of its type of an Euler sum with a distinct MZV and we have no idea why it is true. Any similar MZV proof has been both highly non-trivial and illuminating. To illustrate how far we are from proof: can just the case $n = 2$ be proven *symbolically* as has been the case for $n = 1$? ■

This identity was discovered by the British quantum field theorist David Broadhurst and me during a large hunt for such objects in the mid-nineties. In this process we discovered and proved many lovely results (see [Borwein/Bailey 2003], Chapter 2, and [Borwein et al. 2004], Chapter 4), thereby illustrating #1,#2, #4, #5 and #7. In the case of 'identity' (6) we have failed with

[29] A quite subtle proof has now been found by Zhao and is described in the second edition of [Borwein/Bailey 2003].

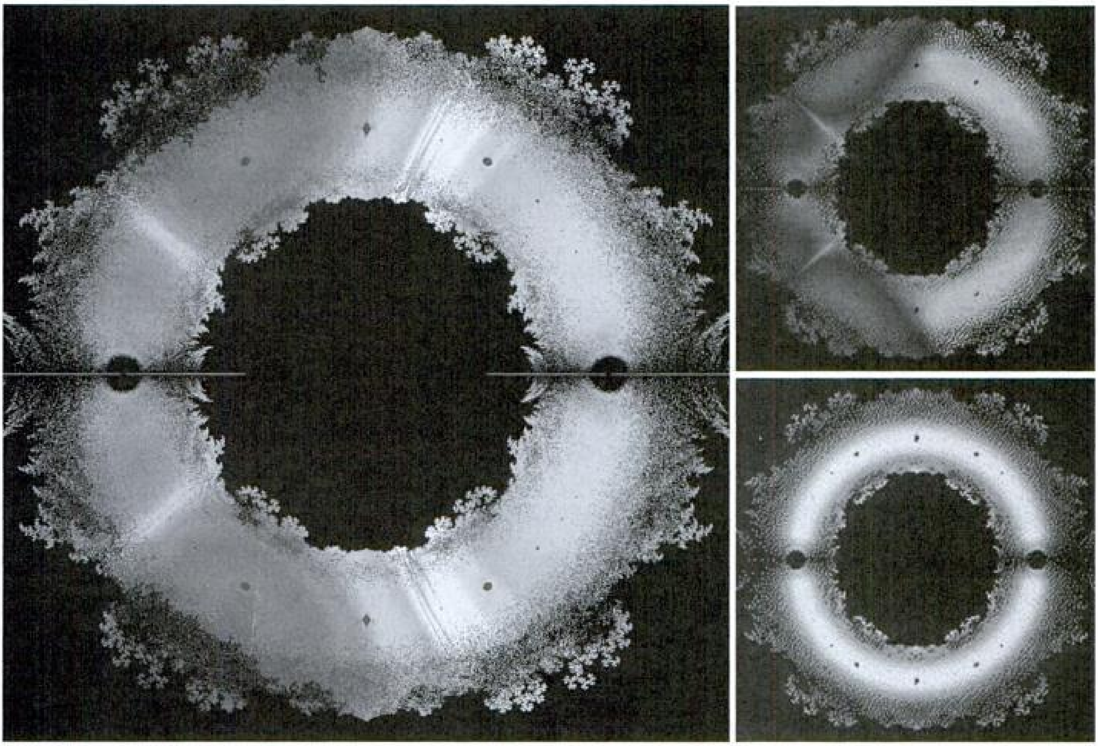


Figure 2.2 **(Ex. 6.): "The price of metaphor is eternal vigilance."** (Arturo Rosenblueth & Norbert Wiener, [Lewontin 2001])

#6, but we have ruled out many sterile approaches. It is one of many examples where we can now have (near) certainty without proof. Another was shown in equation (1) above.

6. **What you draw *is* what you see.** *Roots of polynomials with coefficients 1 or* -1 *up to degree 18.*

   As the quote suggests, pictures are highly metaphorical. The shading in Figure 2.2 is determined by a normalized sensitivity of the coefficients of the polynomials to slight variations around the values of the zeros with red indicating low sensitivity and violet indicating high sensitivity.[30] It is hard to see how the structure revealed in the pictures above[31] would be seen other than through graphically data-mining. Note the different shapes—now proven by P. Borwein and colleagues—of the holes around the various roots of unity.

   The striations are unexplained but all re-computations expose them! And the fractal structure is provably there. Nonetheless different ways of measuring the stability of the calculations reveal somewhat different features. This is very much analogous to a chemist discovering an unexplained but robust spectral line. ■

This certainly illustrates #2 and #7, but also #1 and #3.

[30] Colour versions may be seen at `oldweb.cecm.sfu.ca/personal/loki/Projects/Roots/Book/` and on the cover of this book.

[31] We plot all complex zeroes of polynomials with only -1 and 1 as coefficients up to a given degree. As the degree increases some of the holes fill in—at different rates.

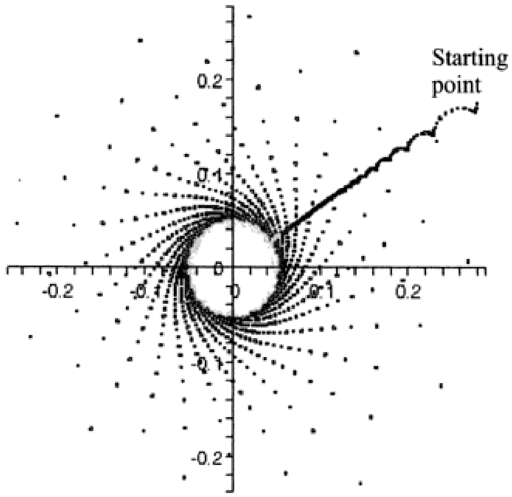


Figure 2.3 **(Ex. 7.): "Visual convergence in the complex plane"**

7. **Visual Dynamics.** In recent continued fraction work, Crandall and I needed to study the *dynamical system* $t_0 := t_1 := 1$:

$$t_n := \frac{1}{n} t_{n-1} + \omega_{n-1} \left(1 - \frac{1}{n}\right) t_{n-2},$$

where $\omega_n = a^2, b^2$ for n even, odd respectively, are two unit vectors. Think of this as a **black box** which we wish to examine scientifically. Numerically, all one *sees* is $t_n \to 0$ slowly. Pictorially, in Figure 2.3, we *learn* significantly more.[32] If the iterates are plotted with colour changing after every few hundred iterates,[33] it is clear that they spiral roman-candle like in to the origin:

Scaling by $\sqrt{n}$, and distinguishing even and odd iterates, *fine structure* appears in Figure 2.4. We now observe, predict and validate that the outcomes depend on whether or not one or both of a and b are roots of unity (that is, rational multiples of π). Input a pth root of unity and out come p spirals, input a non-root of unity and we see a circle. ■

This forceably illustrates role #2 but also roles #1, #3, #4. It took my coauthors and me, over a year and 100 pages to convert this intuition into a rigorous formal proof, [Bailey/Borwein 2005]. Indeed, the results are technical and delicate enough that I have more faith in the facts than in the finished argument. In this sentiment, I am not entirely alone.

Carl Friedrich Gauss, who drew (carefully) and computed a great deal, is said to have noted, *I **have** the result, but I do not yet know how to get it.*[34] An excited young Gauss writes: "A new field of analysis has appeared to us, self-evidently, in the study of functions etc." (October 1798,

---

[32] ... "Then felt I like a watcher of the skies, when a new planet swims into his ken." From John Keats (1795–1821) poem *On first looking into Chapman's Homer*.

[33] A colour version may be seen on the cover of [Bailey et al. 2007].

[34] Like so many attributions, the quote has so far escaped exact isolation!

Proof and Other Dilemmas: Mathematics and Philosophy

Bonnie Gold & Roger A. Simons, Editors

Proof and Other Dilemmas contains sixteen original essays, in language accessible to mathematicians and mathematics students, sketching positions on a variety of contemporary issues in the philosophy of mathematics. Has the use of computers changed the nature of mathematical knowledge? Should it? Is there an empirical aspect to mathematics after all? Are there mathematical objects, and how can we come to know them? Is mathematics socially constructed? Is mathematics the "science of patterns"? In each case, a connection is made to the teaching of mathematics.

HERE ARE EXCERPTS FROM SEVERAL OF THE CHAPTERS:

One can't do mathematics for more than ten minutes without grappling, in some way or other, with the slippery notion of equality. Slippery, because the way in which objects are presented to us hardly ever, perhaps never, immediately tells us...when two of them are to be considered equal.... The heart and soul of much mathematics consists of the fact that the 'same' object can be presented to us in different ways.

—From Barry Mazur's chapter

I believe that more-and-more the interesting parts of mathematics will be less-and-less susceptible to classical deductive analysis...

—From Jonathan Borwein's chapter

Proof as Exploration: Every mathematician knows that when he/she writes out a proof, new insights, ideas, and questions emerge. Moreover, the proof requires techniques, which may then be applied to the consideration of new problems. What makes this topic interesting, and somewhat complex, is that there is not always a hard line between explanation and exploration.

—From Joseph Auslander's chapter

... Mathematics can more appropriately be regarded as a particular kind of study of structures: one that does not imply the existence of special mathematical objects.

—From Charles Chihara's chapter

Unconscious mathematizing is a huge mystery. Mathematical ideas can come into consciousness by surprise, as if by a gift from nowhere - evidently from some subconscious process.

—From Reuben Hersh's chapter

ISBN 978-0-88385-567-6

9 780883 855676