2475/3877

Errata

p 12, line 8: "attempt to construct" for "attempt construct"
p 26, line 12: "confers on its parts" for "confers on it parts"
p 27, line 5: "causal contact" for "casual contact"
p 45, line 1: "Instead he adopts" for "Instead adopts"; line 10: "know that in" for "know that it in"
p 54, line 7: "there are few" for "there a few"
p 56, line 5: "unobtainable" for "obtainable"; line 7: "adopting the idea" for "adopting idea"
p 63, para 3, line 5: delete full stop after "formula A(S,o,N)"
p 72, para 2, line 6: "provided" for "provide"
p 92, line 8: "causal" for "casual"
p 100, lines 8 and 19; p 101, line 1: "canons" for "cannons"
p 140, line 13: "tell" for "tells"
p 218, line 11: "to appeal to" for "to appeal"
p 224, line 10: "versions" for "version"
p 267, line 15: "with the axiom" for "the axiom"
p 320, para 3, line 4: "the evidence for that axiom" for "the evidence for it"; "That is," for "That, is"
p 345, line 13: "based on two" for "based two"
p. 349, para 3, line 2: "*why is the Pythagorean theorem true?*" for "*why is the Pythagorean theorem is true?*"
p 359, line 8: "former" for "latter"
p 380, footnote 1: "of our everyday beliefs" for "of everyday beliefs"
p 381, para 2, line 2: "for mathematics" for "for a mathematics"
p 384, line 7: "… is that the theory" for "… that the theory"
p 386, footnote 8: [Salmon 1989] for [Salmon 19xx]
p 389, line 10: "became incorporated" for "becomes incorporated"; "In the same way … " for "In the same …"
p 389, para 3, line 4: "branches" for "braches"
p 391, line 3: "from" for "form"; line 5: "sciences" for "science"; line 7: "from which" for "from"; para 2, line 8: "meaningful mathematical propositions' for "meaningful propositions"
p 392, line 3: delete "on the other hand"; para 3, line 4: "our" for "out"

Evidence and Explanation in Mathematics

Samuel John Butchart

B.A.(London) M. Phil. (Cantab.)

Submitted for the Degree of Doctor of Philosophy

Department of Philosophy

Monash University

25th May 2001

CONTENTS

ABSTRACT

This thesis is concerned with some problems in the epistemology of mathematics. In the first three chapters, I argue that the proper approach to this subject is *descriptive* rather than *normative*. The aim of the epistemology of mathematics should be to present an account of the ways in which mathematical beliefs are justified which illuminates the practice, history and methodology of mathematics. In the second half, I develop an account of mathematical evidence based on the concept of explanatory unification. I argue that the account of mathematical evidence developed here reveals that despite appearances to the contrary, mathematics is a science like any other.

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university and to the best of my knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Signed: [redacted]

Sam Butchart

ACKNOWLEDGEMENTS

I would like to thank my supervisor, John Bigelow, for his help, encouragement and enthusiasm and my parents, for supporting me, especially during the last six months of the preparation of this thesis. I owe a debt to Bill Hart who introduced to me the philosophy of mathematics and to Marcus Giaquinto, from whom I learned a great deal. I would also like to thank the following people; Lloyd Humberstone, Su Rogerson, Steve Curry, Neil Levy, Neil McKinnon, Konrad Talmont-Kaminski, Brian Weatherson, Steve Curry, Tracy Horn, my brothers Joe and Ben, Emma Lincoln, Sophie, Lucy and Gary, Trevor Luxford, Pete Butland, Kylie McShane, Ross Kirkman and Kim Little. This thesis is dedicated to my parents, with love and gratitude.

INTRODUCTION

There are two main kinds of philosophical question that can be asked about mathematics; ontological questions and epistemological questions. Ontological questions are concerned with the subject matter of mathematics. What, if anything, is mathematics about? An obvious kind of answer immediately suggests itself. The different branches of mathematics are about different things. Number theory is about the natural numbers, geometry is about points, lines and shapes, set theory is about sets and so on.

This kind of answer, of course, immediately raises further questions. What kinds of objects are numbers, points, lines and sets? Are they, as Plato believed, objectively existing, but non-physical or abstract objects, or are they, as others have thought, mental constructions or ideas? Or are they physical objects? Perhaps it is a mistake to think of the subject matter of mathematics in terms of particular objects. Perhaps the subject matter of mathematics concerns universals; properties or relations of some kind. It is often said, for example, that mathematics is in general the study of patterns or structures and that the different branches of mathematics study different kinds of pattern. Then again, perhaps the question with which we began is based on a false presupposition. According to formalists, for example, mathematics has no subject matter. On this view, mathematics is essentially a meaningless game with symbols, played according to arbitrary rules, like chess.

On the other hand, we have epistemological questions; questions concerning our knowledge of mathematics. How do we come to know the truths of mathematics? How is mathematical knowledge acquired? Plato held that mathematical knowledge was in some sense innate. When we come to know a mathematical truth, we are really just remembering something we knew already but had forgotten.[1] Later platonists held that mathematical knowledge was not innate but learned, via a faculty of mathematical intuition, akin to, but distinct from sensory perception, which gives us access to the platonic realm of abstract objects.[2] Others have held that mathematical knowledge is ultimately derived from the senses. J. S Mill, for example, argued that arithmetical statements such as '2 + 2 = 4' were simply highly confirmed inductive generalisations, gained from our experience of counting collections of physical objects.[3] Again, perhaps the whole question is based on a mistake. If, as the formalists say, mathematics is simply a game we play with symbols, then there is a sense in which we have no mathematical knowledge. If mathematics has no subject matter, there is no question of how we can acquire knowledge of the objects of mathematics. Perhaps all there is to mathematical knowledge is knowledge of technique, 'know how' rather than 'knowing that', like knowing how to ride a bike or how to get mate in three moves from a certain chess position.

Answers to the question of how we acquire mathematical knowledge suggest further epistemological questions. What is the epistemic status of the knowledge so acquired? How does it differ, if at all, from other kinds of human knowledge? Is mathematical knowledge certain and infallible? Or is it, like our knowledge of physics and chemistry, uncertain and fallible, open to revision in the light of new evidence?

Obviously these epistemological and ontological problems are interdependent. What, if anything, you take mathematics to be about will affect how, if at all, you think mathematical knowledge is acquired. This is most clearly apparent in the case of formalism; if you hold that mathematics is not a body of propositions which are about

[1] This doctrine is described in the *Meno*. See [Jowett 1949].
[2] See for example [Gödel 1947].
[3] See [Mill 1884].

anything at all, you are forced to the conclusion that there can be no mathematical knowledge.[4] Likewise, Plato's account of mathematical knowledge was dictated by his account of the subject matter of mathematics as concerning a realm of abstract objects. Such objects, being non-physical cannot be perceived and hence, Plato argued, our knowledge of them cannot be based on the evidence of the senses. The only alternative Plato saw was to say that our mathematical knowledge was innate.

Conversely, the way in which you think mathematical knowledge is acquired will affect what you take the subject matter of mathematics to be. Mill, for example, as an empiricist, held that all human knowledge must ultimately derive from the senses. In particular, this has to be true for mathematical knowledge. But if mathematical knowledge is empirical, it seems that it cannot be knowledge of abstract, non-physical objects which are inaccessible to perception. Therefore, mathematics must be about physical objects or processes of some kind. Arithmetic, for example is a body of truths about the results of counting and operating on collections of physical objects.[5]

Clearly, we would like to be able to answer both kinds of question and we would like our answers to be consistent with each other. The main difficulty in the philosophy of mathematics has not been that we have failed to find consistent answers to these questions. Rather, the problem has always been one of finding answers to the two kinds of questions

[4] One might say that we can have knowledge *about* the formalism itself, knowledge of the consistency of a certain formal system for example. But as Frege argued, either meta-mathematics is mathematics, in which case not all mathematics is simply a formal game and some of it is factual after all, or meta-mathematics itself is open to a formalist description, in which case, we have no factual mathematical knowledge. See [Frege 1884, §§93-119, Dummett 1991, pp. 253-5].
[5] This is an example of the famous saying 'One philosopher's *modus ponens* is anothers *modus tollens*.' Both Plato and Mill accept the conditional; if mathematics is about abstract objects, it cannot be known via the evidence of the senses. Plato accepts the antecedent and applies *modus ponens*, arriving at the doctrine of *anamnesis*, Mill rejects the consequent, applies *modus tollens* and arrives at, what Frege derisively called his 'pebble arithmetic'. In more recent times, Quine has famously denied the conditional, arguing that there can be *empirical* evidence for *abstract* objects.

which are simultaneously acceptable. Plato's answers for example, are consistent with each other, but not simultaneously acceptable. The claim that mathematics is about abstract objects has much to recommend it; the doctrine that mathematical knowledge is innate does not. In general, platonism, while independently plausible as an account of the ontology of mathematics faces serious epistemological difficulties. Plato's solution to those difficulties would not now have many adherents, but more seriously, as we shall see, some have argued that no other solution is possible in principle. Likewise, Mill's empiricism is independently plausible, but his account of the subject matter of mathematics is open to severe objections.[6]

The final aim of the philosophy of mathematics should be to provide answers to the ontological and epistemological questions about mathematics which are not only consistent with each other but simultaneously acceptable. Although much progress has been made, we are still a long way from attaining this goal. What makes it hard to arrive at simultaneously acceptable answers to all these questions are the complex interconnections between them. It is as though we are faced with a tangled knot; pulling out the loops in one place serves only to tighten up the knots in another.

In this thesis, I will be concerned with epistemological questions, rather than ontological ones. One aim is to show that there are many questions and problems in the epistemology of mathematics which are quite independent of any particular account of its subject matter. Some examples: Can there be empirical evidence for mathematical statements? How does a mathematical proof provide us with evidence for its conclusion? How are the axioms and other first principles of mathematical theories justified? Is

[6] See, for example, [Frege 1884, §§7-11, Dummett 1991, pp. 58-61, Ayer 1936]. For a reconstruction and defence of Mill's views see [Kitcher 1980].

deductive justification the only kind of mathematical evidence? How can we make sense of the idea, common to many mathematicians, that some proofs of a theorem show not only *that* it is true, but also explain *why* it is true? I will attempt to provide answers to all these questions in what follows. A second aim is to argue that *mathematics is a science*. Mathematics is a science, not because it is empirical, but because it is sensitive to *evidence*. The central problem I will be concerned with then, is that of describing and explaining the nature of evidence in mathematics.

One kind of view concerning the nature of evidence in mathematics is *foundationalism*. On this view, the proper justification of our mathematical beliefs consists in showing how they can all be derived from a set of epistemologically secure first principles. I critically examine this idea in chapter one, taking Frege's logicist programme as my example.

In epistemological terms, Frege's work represents two things. Firstly, it is an example of a foundational account of mathematics, a description of the kind of evidence we have for mathematics. I argue that as such it is a failure; mathematics is not built on unique epistemologically privileged foundations. Secondly, Frege's work is an example of a certain kind of approach to epistemological questions; an approach I want to call *normative*, as opposed to *descriptive*. On this approach, the task of the epistemologist is to attempt to *justify* mathematics, to provide reasons for believing that mathematics is *true*. Foundationalism is the view that we can do this by delineating a set of first principles in terms of which the justification can be carried out. Although the foundational ideal has been largely abandoned in recent philosophy of a mathematics,[7] the epistemological problem is still often seen as one of providing a justification for mathematics. But why should mathematics need justifying if we have given up the idea that it needs some kind of *a priori* foundation? The feeling that there is still a problem of justification here stems from the fact that mathematics appears to be concerned with abstract, non-physical objects. The problem is then to explain how belief in such objects can be justified, given a broadly naturalistic or empiricist account of human knowledge in general. One popular kind of approach to this problem is to attempt construct a non-platonist ontology for mathematics which allows for an empirical or perceptual route to knowledge of mathematics. In chapter two, I examine some of this recent work and suggest some grounds for dissatisfaction with the normative approach. This leads, in chapter three, to a more detailed examination of an alternative, descriptive approach to the epistemology of mathematics. Here I focus on the work of Imré Lakatos and Philip Kitcher. This serves not only to illustrate the kind of descriptive approach to epistemological questions that I will be taking here, but also to elaborate the thesis that mathematics is a science, by showing how its growth and development is sensitive, in recognizably scientific ways, to evidence of various kinds.

I begin my investigation of the varieties of mathematical evidence with a discussion of empirical evidence. Quine has famously argued that since mathematics is an integral part of our best scientific theories and since those theories have been successfully confirmed by experiment and observation, it follows that there is empirical evidence for mathematics. I examine Quine's argument in chapter four. I argue that although there are some serious problems with Quine's argument; problems connected with his general account of scientific

[7] There are exceptions of course. Most notably, Crispin Wright has attempted a revival of Frege's foundationalist programme, while Michael Dummett has argued for a version of intuitionism. See [Wright 1983, Dummett 1975].

evidence; it is nonetheless indisputable that empirical evidence does play *some* role in mathematics. Where I part company with Quine is on the question of whether empirical evidence is the *only* kind of evidence we can have for mathematics. In fact, although utility in science is one source of evidence in mathematics, it is not the only source, or even the main source. Far more important than utility in science is utility in *mathematics*.

In chapter five I take a closer look at this kind of mathematical evidence and sketch the outlines of a general account. I argue that the key to giving an adequate account of mathematical evidence lies in understanding the concept of *explanation* in mathematics. In the final chapter, I attempt to develop an account of mathematical explanation and show how it can cast some light on the nature of evidence in mathematics. I conclude with some remarks on how this investigation of mathematical evidence reveals that despite appearances to the contrary, mathematics is a science like any other.

CHAPTER ONE

FOUNDATIONAL EPISTEMOLOGY

Many philosophers have supposed that the epistemological problem in the philosophy of mathematics is to show how mathematical knowledge is possible by providing a justification for our mathematical beliefs. An example of this kind of approach is the view (or family of views) I call *foundationalism*. The central idea here is that all of our mathematical beliefs can be justified in terms of an epistemologically privileged set of first principles. The task of the philosopher is to uncover these first principles or foundations and show how the required justification can be carried out.

In this chapter, I want to examine and criticise this idea. I will take as my example the logicist programme of Gottlob Frege, not only because it represents the first fully worked out attempt to provide mathematics with a foundation, but also because it raise many important epistemological issues which will occupy us in later chapters. I describe Frege's strategy for showing that arithmetic is a branch of logic and explain why that strategy failed. I then examine some of Frege's epistemological assumptions. We can distinguish in Frege's work, two distinct senses in which a set of first principles may be said to provide a foundation for mathematics; they may provide us with *certainty* regarding mathematical truths or they may reveal the *proper justification* for those truths. I argue that any attempt to provide mathematics with foundations in either sense is misguided. Mathematics, like any other science, does not have foundations. There are no unique, epistemologically privileged first principles in terms of which all our mathematical beliefs

receive their proper justification. I will begin however, by reviewing some of the philosophical and mathematical motivations for Frege's programme.

1. FREGE'S PROGRAMME

In 1884, Frege published what has come to be seen as one of the great classics of philosophy, *Die Grundlagen Der Arithmetik.* Frege begins the introduction to the book by remarking that no satisfactory answer to such an apparently simple question as 'what is the number one?' has yet been given, nor has anyone been able to answer the more general question 'what is a number?'. Frege describes this situation as a scandal to the science of mathematics:

> Questions like these catch even mathematicians for that matter, or most of them, unprepared with any satisfactory answer. Yet is it not a scandal that our science should be so unclear about the first and foremost among its objects, and one which is apparently so simple? Small hope, then that we shall be able to say what number is. If a concept fundamental to a mighty science gives rise to difficulties, then it is surely an imperative task to investigate it more closely until those difficulties are overcome; especially as we shall hardly succeed in finally clearing up negative numbers, or fractional or complex numbers, so long as our insight into the foundation of the whole structure of arithmetic is defective.
>
> [Frege 1884, p. II]

An analysis of the concept of number is therefore required if arithmetic is to retain its status as a science. The motivation for Frege's work is set out in more detail in §§1 and 2 of *Grundlagen.* In §1 Frege describes his project as in line with contemporary attempts to instil rigour in analysis. By the beginning of the nineteenth century, many mathematicians had become dissatisfied both with the clarity of the basic concepts of analysis (those of function, infinite series, continuity, derivative and so on) and with the rigour of proofs which had been given of the theorems of analysis. For example, consider a function which is continuous over the interval [a, b]. Suppose this function has a negative value at a and a positive value at b. If we imagine the graph of the function, it is obvious that the function must cross the x-axis at some point in the interval [a, b]. Many mathematicians appealed to such considerations in order to 'prove' that any such function must take the value zero at some point in the interval. Here a fundamental theorem of analysis is justified by appeal to geometric intuition.

Bolzano was one of the first mathematicians to argue for the elimination of such appeals to intuition from analysis. The need to do this was felt to be especially pressing in the case of analysis, for here our geometric intuition can easily lead us into error. It had been intuitively obvious to many mathematicians that any continuous function must be differentiable, except perhaps at a finite number of isolated points. Bolzano himself found an example of a function which is continuous but *nowhere* differentiable. In order to attain some degree of certainty regarding the theorems of analysis then, we ought to avoid geometric intuition. Bolzano also believed that analysis was a science independent of geometry - its true subject matter was not points and curves, but *quantities* - real numbers and functions. If so, the real grounds for accepting the theorems of analysis cannot be geometrical, they must in some sense by arithmetical.

Bolzano began then, the task of reducing the concepts of analysis to purely arithmetical concepts. This would achieve two things: appeals to uncertain geometric intuition would be eliminated from proofs of theorems and the true basis of those theorems would be exhibited. Bolzano attempted to give a 'purely analytic' proof of the theorem

mentioned above - the intermediate value theorem.[1] This work was continued and improved upon by such mathematicians as Cauchy and Weierstrass.[2]

Frege wished to achieve similar results for arithmetic - the theory of the natural numbers 0, 1, 2, ... Here too, many things are accepted without rigorous proof, by appeal to induction or intuition; Frege mentions for example, such numerical formula as 7 + 5 = 12 or the associative law of addition. These, says Frege. "... are so amply established by the countless applications made of them every day, that it may seem almost ridiculous to try to bring them into dispute by demanding a proof of them. But it is in the nature of mathematics always to prefer proof, where proof is possible, to any confirmation by induction." [Frege 1884, §2]. The nineteenth century mathematicians had shown that obvious and fundamental truths of analysis could in fact be proved. Frege hoped to achieve the same for arithmetic. Not only would we thereby achieve greater certainty, by eliminating appeals to uncertain intuitions, we would also attain "insight into the dependence of truths upon one another" [ibid. §2]. That is, the true foundations of arithmetic would be revealed.[3]

These then, were the mathematical motivations for Frege's work. In §3, he discusses a philosophical motivation. Since Kant, philosophers had debated whether mathematics was *a priori*, *a posteriori*, analytic or synthetic. Frege now gives his own definitions of these philosophical terms. For Frege, these are distinctions among kinds of

---

[1] In his 'Rein analytische Beweis' [Bolzano 1817]. See also [Kitcher 1975] for a philosophical discussion of Bolzano's proof.

[2] See for example [Kline 1972, pp. 947-978] and [Kitcher 1984, pp. 229-71]. We shall look at some of this work in more detail in chapters three and five.

[3] As we shall see in chapter three, Philip Kitcher has argued that Frege was mistaken in comparing the mathematical motivation for his programme with the attempts to instil rigour in analysis. In the latter case, the vagueness of the fundamental concepts was becoming a severe impediment to solving the problems mathematicians were interested in; no such state of affairs was impeding the progress of number theory. The concepts of analysis were indeed unclear and problematic, but it was not obvious that the concept of natural number was in urgent need of clarification and analysis. [Kitcher 1979, p. 239; 1984, pp. 268-270].

true judgement and concern "not the content of the judgement but the justification for making the judgement." [ibid. §3]. According to Frege, such a justification must be a deductive proof of the proposition in question from primitive truths. The status of the proposition then depends on the nature of the primitive truths appealed to in its proof:

> The problem becomes, in fact, that of finding the proof of the proposition, and of following it up right back to the primitive truths. If, in carrying out this process, we come only on general logical laws and on definitions, then the truth is an analytic one, bearing in mind that we must take account also of all propositions upon which the admissibility of any of the definitions depends. If, however, it is impossible to give the proof without making use of truths which are not of a general logical nature, but belong to the sphere of some special science, then the proposition is a synthetic one. For a truth to be a posteriori it must be impossible to construct a proof of it without including an appeal to facts, i.e truths which cannot be proved and are not general, since they contain assertions about particular objects. But if, on the contrary, its proof can be derived exclusively from general laws, which themselves neither need nor admit of proof, then the truth is a priori.
>
> [ibid. §3]

Thus true judgements are divided into two main categories; a truth is *a posteriori* if *every* proof of it appeals to at least one fact, where a fact is an unprovable judgement of the form *Fa*; a predication of a property to a particular object. A truth is *a priori* if it is not *a posteriori*; that is if *some* proof of it appeals only to general laws.[4] Among *a priori* truths Frege further distinguishes between analytic and synthetic judgements, depending on the nature of the general laws appealed to in the proof of the judgement. If all of the laws are "of a general logical nature", then the judgement is analytic. If the proof appeals to non-logical laws, belonging to "the sphere of some special science", then the judgement is synthetic *a priori*.

[4] Frege seems to be assuming here that every proposition is either a general law or states a particular fact. The assumption might be doubted; existentially quantified statements like 'Whales exist' are not clearly general laws, nor are they predications of a property to a particular object.

The following diagram summarises Frege's distinctions.

True Judgements

A Posteriori
<u>All</u> proofs appeal to
at least one fact.

A Priori
<u>Some</u> proof appeals
<u>only</u> to general laws.

Analytic
<u>Only</u> logical laws
appealed to

Synthetic
<u>Some</u> non-logical
laws appealed to

For Frege, the distinction between general logical laws and non-logical laws is that the non-logical laws contain terms which refer to some specific domain of discourse, while logical laws consist only of terms which are universally applicable. Logical laws are thus 'topic neutral'. The axioms of geometry were for Frege of the first kind; they contain terms which are not of universal application, but apply only to things which can be said to exist in space. Hence, Frege agrees with Kant that geometry is synthetic *a priori*.

However, Frege disagreed with Kant about the status of arithmetic. Arithmetic is not synthetic *a priori*, as Kant had held, but analytic.[5] To establish the analyticity of arithmetic then, Frege needs to show that the propositions of arithmetic can be proved from logical laws alone. This is the task he sets himself in *Grundlagen*; to show that the basic laws of arithmetic can be deduced from purely logical first principles. If every truth of

[5] It should be clear that Frege's definition of 'analytic' is utterly at odds with the version of this concept familiar from the writings of the logical positivists. The latter held that analytic truths do not state 'matters of fact' but simply record the conventions we have adopted for the use of certain symbols; analytic statements are true in virtue of the meanings of their component terms. On this account, the claim that arithmetic is analytic is meant to be inconsistent with the view that arithmetical truth requires the existence of abstract objects; nor does it entail that arithmetic is derivable from the laws of logic. [See Ayer 1936]. Frege's view, by contrast, is certainly *not* that analytic statements are made true by linguistic convention; in particular, arithmetical truths are not only derivable from the laws of logic, they also state facts about certain abstract objects – as we shall see, Frege combines his logicism with a thoroughgoing *platonism*.

arithmetic is a deductive consequence of these basic laws, then Frege will have shown that arithmetic is, after all, analytic and neither synthetic *a priori*, as Kant had held or *a posteriori*, as J. S. Mill had suggested.

It is useful to distinguish three components of Frege's logicism:

(1) Arithmetical concepts can be defined using purely logical vocabulary.

(2) The basic laws of arithmetic can be deduced from purely logical laws, using those definitions.

(3) Every truth of arithmetic is analytic; deducible from purely logical laws by means of definitions.

These claims are independent of each other. It is now well known that (2) does not entail (3), since arithmetic cannot be given a complete axiomatisation. Furthermore, the claim that arithmetical concepts can be expressed in purely logical vocabulary (1) and the claim that the basic laws or axioms of arithmetic are analytic (2) are logically independent. For example, the Dedekind-Peano axioms for number theory (see below) contain the apparently non-logical symbols *zero*, *successor* and *natural number*. But since it might be possible to give definitions of these symbols in purely logical terms (and this is what Frege sets out to do), this does not by itself prevent those axioms from being analytic. Conversely, a non-analytic proposition might be expressible in purely logical terms. Consider a claim like 'There are at least two individuals', which we can symbolise as $\exists x \exists y (x \neq y)$. This is a formula expressed in purely logical terms, but which is not a consequence of the laws of logic and so not analytic.

What are the basic laws of arithmetic? Richard Dedekind addressed this question in his book *Was sind und was sollen die Zahlen?*, written at about the same time as *Grundlagen*. The Preface to the first edition of Dedekind's book, reveals that he had much in common with Frege:

> In science nothing capable of proof ought to be accepted without proof. Though this demand seems so reasonable yet I cannot regard it as having been met even in the most recent methods of laying the foundations of the simplest science; viz., that part of logic which deals with the theory of numbers. In speaking of arithmetic (algebra, analysis) as a part of logic I mean to imply that I consider the number concept entirely independent of the notions or intuitions of space and time, that I consider it an immediate result from the laws of thought.
>
> [Dedekind 1888, p. 31]

So far, Frege would agree entirely. But Dedekind continues:

> My answer to the problems propounded in the title of this paper is, then, briefly this: numbers are free creations of the human mind; they serve as a means of apprehending more easily and more sharply the differences of things.
>
> [Dedekind, *op cit.*]

This Frege would certainly not have agreed with, being firmly opposed to any such 'psychologistic' account of arithmetic. He held that mental processes such as 'abstraction' which Dedekind (and Husserl) appealed to, were irrelevant to mathematics. Frege's logicism was combined with a thoroughgoing *platonism*; for him, numbers are not mental constructions of any kind, but objectively existing, abstract objects.

In his book, Dedekind does not state explicit axioms for number theory. The first axiomatization of arithmetic was published by Peano[6], although the Peano axioms were in

[6] See [Peano 1889].

fact first stated by Dedekind in correspondence.[7] The Dedekind-Peano axioms for arithmetic can be stated as follows:

(1) Zero is a natural number

(2) Zero is not the successor of any natural number.

(3) Every natural number has a successor, which is also a natural number.

(4) No distinct natural numbers have the same successor.

(5) Any property which:

(a) belongs to zero

(b) if it belongs to a natural number, it also belongs its successor

belongs to every natural number.

The *successor* of a number is just the next number in the series; two is the successor of one, three is the successor of two and so on. The fifth axiom is the principle of mathematical induction. Poincaré held that it was not reducible to logic, but was a specially mathematical, synthetic proposition.[8]

Instead of taking *zero*, *successor* and *natural number* as undefined, Frege sets out to define them, using purely logical concepts. This would satisfy the first component of his logicism; (1) above. Frege then attempts, in effect, to derive the five Dedekind-Peano axioms from these definitions. This would satisfy the second component of his logicism; (2). If every truth of arithmetic were deducible from these axioms, then the third component (3) would also be satisfied.

[7] See [Wang 1957].
[8] See [Poincaré 1894].

This then, was Frege's programme; to show that arithmetic was analytic by giving definitions of the basic arithmetical concepts in purely logical terms and deriving the basic laws of arithmetic from those definitions. The motivation was essentially epistemological. Frege aimed to exhibit the true justification for arithmetic. In particular, he aimed to show that we do not need to appeal to either intuition or perception in order to justify arithmetical propositions.

Grundlagen Der Arithmetik represents only an outline of Frege's programme however, because it contains only informal proof sketches, not fully formal proofs. For Frege, the advantage of formalising proofs was that it would make explicit every assumption, no matter how obvious, that was required in order for a proof to be valid. By formalising the proofs of the basic laws of arithmetic, the assumptions necessary to establish them would be clearly revealed. Informal inferential steps that seemed to appeal to intuition would be revealed to involve no more than a possibly quite long and complex sequence of simple deductive steps. If on the other hand, we did need to appeal to synthetic propositions (and hence to intuition) at some stage in the proof, formalisation would reveal this. Frege had already taken a first step, by inventing one of the first serious formal systems, in his *Begriffsschrift* [1879]. The task of setting out the proofs sketched in the *Grundlagen* in a fully formal setting was carried out by Frege in his *Grundgesetze der Arithmetik* [1893-1903].

The discussion so far should already have revealed some important assumptions in Frege's epistemology; that there is such a thing as the proper justification of any given truth; that this justification is always a deductive proof; that a true science requires clear and precise concepts and that logic provides us with sure and certain knowledge, independent of perception and intuition. I will examine these assumptions in more detail in sections four and five. Before doing so however, I will describe the strategy Frege develops for establishing the analyticity of arithmetic and explain how that strategy led to the disaster of the contradiction in Frege's system.

2. FREGE'S METHOD

§§5-61 of *Grundlagen* are concerned with the question 'what is a number?'. More accurately, Frege wants to describe the logical form of *ascriptions of number* – propositions such as 'two horses are pulling the carriage' or 'there are five books on the table'. Frege rejects the view that here we are ascribing a property – 'fiveness' – to the books on the table; number is not a property of objects. He arrives at the conclusion that an ascription of number assigns a property, not to objects, but to concepts. A concept is something which can be either true or false of particular objects. For example, we have the concept 'is a book on the table'. When we say that there are five books on the table, we are ascribing a property to this concept, the property of having five instances, or in Frege's terminology of there being five objects that 'fall under' the concept.

Notice that this does not tell us what sort of thing a number is, it just tells us that ascriptions of number predicate something of a concept. In §§56-61, Frege argues for the view that numbers are objects. In particular, they are non-physical, non-mental, yet objective objects; abstract objects in other words.[9] This is the platonist strand in Frege's

[9] Frege calls them *non-actual* objects. This does not mean that numbers are merely *possible* objects - objects which do not actually exist, but might; all Frege means is that numbers are real, but non-physical, non-causally active objects.

logicism. If numbers are objects, then an ascription of number should contain a singular term or name for such an object. Frege argues, that in fact, an ascription of number like:

(1) There are n Fs

has the logical form:

(2) The number belonging to the concept F = n

where n is a singular term denoting a particular number. Suppose we write Num(F) for 'the number belonging to the concept F'. Then Frege's claim is that (1) has the form:

(3) $\mathrm{Num}(F) = n$

The symbol Num(F) represents an operator on concepts; a function which takes a concept as argument and returns an object as its value. We will call this operator the *cardinality operator*.

Having argued that numbers are objects, Frege asks the Kantian question 'how are numbers given to us?'; that is, 'how do we acquire knowledge of numbers?'. Kant had said that objects can only be given to us or known through 'sensible intuition', sensory perception in other words. But Frege has by now rejected the view that number is a physical or perceptible property of things or that we can have intuitions of numbers. How then do we acquire knowledge of numbers, how are they 'given' to us, if they cannot be perceived or apprehended by intuition?

This is a general problem for platonism. Abstract objects are not part of the physical universe, they do not exist in space or time. Being non-physical, they are entirely causally isolated from us, they cannot affect us, or any other physical thing in any way. But if abstract objects are causally unreachable in this way, how can we get to know facts about them, or even become aware of their existence at all? The obvious answer is that we cannot.

Hence either we have no mathematical knowledge, or platonism is not the correct view of mathematics.[10]

Frege attempted to solve this problem by appeal to what has since become known as the *context principle*; "only in the context of a proposition do words really have a meaning". [Frege 1884, §60]. According to Frege, the objection to abstract objects just considered is based on a fallacious assumption; that words acquire meaning by being associated with ideas, so that in order to refer to an object of a certain kind, we must have some idea or intuition of it. This is a view of meaning and reference which Frege rejects. For Frege, the meaning of a word is given the systematic contribution it makes to the truth-conditions of sentences. A word has meaning for us if we know how it affects the truth-conditions of all the sentences in which it can occur. "It is enough", Frege says, "if the proposition taken as a whole has a sense; it is this that confers on it parts also their content." [ibid. §60]. Hence, if we can give the truth-conditions for all contexts (sentences) in which singular terms for numbers can occur, then we can justify the claim that numbers are objects, even though we cannot causally interact with, perceive or imagine the objects those words refer to.

In giving the truth-conditions for all contexts in which numerical terms can occur, we have conferred a sense or meaning on those words, but clearly this does not imply that they actually refer to anything. We could give truth-conditions for statements involving unicorns (say by means of some definition of the predicate 'x is a unicorn') but it does not follow from this that there really are any unicorns. However, if some sentences involving the word are actually *true*, then the word must have a reference. So Frege's claim would be that if we can give a sense to numerical terms by laying down conditions for the truth or

[10] I discuss a modern formulation of this ancient problem in the next chapter.

falsity of all sentences containing them (by means of appropriate definitions of those terms), *and* we can establish that certain of those sentences are in fact true (and indeed, Frege hopes to show that they are *logically* true), then it follows that numerical terms have a reference and in particular that they refer to objects, even though we can have no perceptual or casual contact with those objects.[11]

According to the context principle then, we will have shown how reference to and knowledge of numbers is possible, despite their abstractness, if we can give determinate truth-conditions to sentences containing terms for numbers. We do not, in addition, need to show how such objects can be perceived or how we can have intuitions or ideas of them. Frege's answer to the Kantian question then, is that numbers are given to us, not through perception or intuition, but through *language*: "How, then, are numbers to be given to us, if we cannot have any ideas or intuitions of them? Since it is only in the context of a proposition that words have meaning, our problem becomes this: To define the sense of a proposition in which a number word occurs." [Frege 1884, §62]. Michael Dummett argues that Frege's move here represents the first example of the 'linguistic turn' in philosophy.[12]

Frege now argues that since numbers are objects, the primary type of statement which must be supplied with determinate truth conditions are *identity statements*:

> ...we have already settled that number words are to be understood as standing for self-subsistent objects. And that is enough to give us a class of propositions which must have a sense, namely those which express our recognition of a number as the same again. If we are to use the symbol *a* to signify

[11] This interpretation of Frege's use of the context principle is essentially the same as that given by Crispin Wright [1983] and Michael Dummett [1991]. The interpretation of the context principle is a somewhat controversial area, to the say the least. The problem is made all the more difficult because Frege does not distinguish very clearly, in the *Grundlagen*, between an expression having a sense (or meaning) and its having a reference. See [Dummett 1991, pp. 66-7].
[12] See [Dummett 1991, p. 111].

an object, we must have a criterion for deciding in all cases whether b is the same as a, even if it is not always in our power to apply this criterion.

[Frege 1884, §62]

Frege has argued that the fundamental type of numerical term is 'the number of Fs' or, in his terminology, 'the number belonging to the concept F'. His task then, becomes that of stating the truth-conditions for sentences of the form:

(4) The number of Fs = the number of Gs

Only when we have done this can we legitimately say that such terms refer to *objects*. In Quine's phrase; 'no entity without identity'.[13] In §63, Frege suggests a solution to this problem:

Hume long ago mentioned such a means: "When two numbers are so combined as that the one has always a unit answering to every unit of the other, we pronounce them equal." This opinion, that numerical equality or identity must be defined in terms of one-one correlation, seems in recent years to have gained widespread acceptance among mathematicians.

[ibid. §63]

We can distinguish two separate proposals here. The first is that a statement of the form (4); 'the number of Fs = the number of Gs', is equivalent to 'there are just as many Fs as Gs'. 'There are just as many Fs as Gs' in turn is to be defined in terms of the idea of a one-one correlation or mapping: there are just as many Fs as Gs iff there is a one-one correlation between the Fs and the Gs.[14]

[13] See [Quine 1953]. Quine famously used the idea of a criterion of identity to refute the view that there are such things as possible objects. We are unable to state any identity criteria for such objects "and what sense can be found in talking of entities which cannot meaningful be said to be identical with themselves and distinct from one another?" [*ibid.* p. 4].

[14] A relation R provides a one-one correlation between the Fs and the Gs if and only if, every F bears the relation R to exactly one G and conversely, every G is related by R to exactly one F.

Suppose we write $F \approx G$ for 'there are just as many Fs as Gs'; in Frege's terminology, this says 'the concept F is *equinumerous* with the concept G'. We write $\exists R(F\ 1\text{-}1_R\ G)$ to mean there is a relation R, such that R provides a one-one correlation between the Fs and the Gs. Then Frege's proposal can be summarised as follows:

(Hume's Principle) $\mathrm{Num}(F) = \mathrm{Num}(G) \leftrightarrow F \approx G$

(Def. $\approx$) $F \approx G =_{df} \exists R(F\ 1\text{-}1_R\ G)$

Hume's Principle, 'The number of Fs = the number of Gs iff the Fs and the Gs are equinumerous' then provides the required criterion of identity for numbers. In §§63-67 Frege canvasses the possibility of taking Hume's Principle as providing us with a *definition* of the cardinality operator, Num(F). Clearly this would be a *contextual* rather than an *explicit* definition of the operator. An explicit definition has the form either of an identity statement (like the definition of $\approx$ given above) or the form of a logical equivalence; we could write that definition in the form: $F \approx G \leftrightarrow \exists R(F\ 1\text{-}1_R\ G)$. Such a definition allows us the means of replacing the defined term or *definiendum* ($F \approx G$ in this case) in every context in which it occurs, with the defining term or *definiens* (in this case, $\exists R(F\ 1\text{-}1_R\ G)$).

A contextual definition of a term, by contrast, is given by stating truth-conditions only for some contexts in which the defined tenn can occur. For example, we could give a contextual definition of the notion of an ordered pair, by stating the required identity conditions for such objects. That is:

(5) $\langle a, b\rangle = \langle c, d\rangle \leftrightarrow a = c \,\&\, b = d$

This is a contextual definition because we have not defined the symbol $\langle a, b\rangle$ for all contexts in which it can occur, but only for identity contexts. An explicit definition of $\langle a, b\rangle$ could be given in set-theoretic terms by (for example):

(6) $\langle a, b\rangle =_{df} \{\{a\}, \{b, \varnothing\}\}$

from which (5) can be deduced using the standard principles of set theory.

The idea currently under consideration then, is that we take Hume's Principle as a contextual definition of the cardinality operator, Num(F), in much the same way as we could take (6) as a contextual definition of the ordered pair $\langle a, b\rangle$.

In §64, Frege switches the discussion to an analogous case. We could, Frege says, define the concept of *the direction of a straight line* by means of the follow principle:

(7) $\mathrm{Dir}(a) = \mathrm{Dir}(b) \leftrightarrow a \,//\, b$

That is, the direction of a = the direction of b iff a is parallel to b. Hence, given the concepts of *line* and *parallel*, we obtain the abstract concept of a direction. Frege also mentions the concept of shape, which could be introduced in an analogous way:

(8) $\mathrm{Shp}(a) = \mathrm{Shp}(b) \leftrightarrow a$ and b are similar figures.

The shape of a = the shape of b iff a and b are similar.[15] Note the analogy between such definitions and the proposal that we define the cardinality operator by means of Hume's Principle:

(9) $\mathrm{Num(F)} = \mathrm{Num(G)} \leftrightarrow \mathrm{F} \approx \mathrm{G}$

In each case, we define an operator or function, by stating identity conditions for the values of the function, in terms of an equivalence relation (parallelism in the case of directions, similarity for shapes, equinumerosity for numbers) which holds between the arguments of

[15] Similar figures are those such that corresponding sides are in constant proportion.

the function (lines in the case of directions, geometrical figures in the case of shapes, concepts in the case of numbers).[16]

In §§64-65, Frege defends this method of definition against two objections, then in §§66-67, he discusses an objection which he finds compelling and so the proposal is rejected. The objection has become known as the *Julius Caesar Problem*. Essentially, the problem is that such contextual definitions do not uniquely determine the referent of the defined term. The proposed definition of the direction operator (7) for example, can tell us when two directions are identical, but it does not tell us what a direction *is*. It does not tell us, in general, which things are directions and which are not. Hence, given an arbitrary object, like England or Julius Caesar, the definition cannot tell us whether that object is or is not a direction.

The same applies to the proposed contextual definition of the cardinality operator (9). The definition tells us when the numbers belonging to two concepts are identical, but it does not tell us what a number *is* - which things are and which things are not, numbers. The definition fails to tell us the truth-value of propositions of the form Num(F) = q, except where q itself is of the form Num(G) for some concept G.

Why should this matter? Sentences like 'the number three = Julius Caesar' are not likely to be ones we are going to want to either prove or disprove in mathematics. However, recall that according to the context principle, *all* that is necessary in order to fix the meaning of a term, is to specify truth-conditions for sentences in which it can occur. It is not necessary, for example, to explain how we can have ideas, intuitions or causal

[16] An equivalence relation is one that is reflexive, symmetric and transitive. That is, R is an equivalence relation iff:

(1) $\forall x(\mathrm{R}xx)$

(2) $\forall x\forall y(\mathrm{R}xy \rightarrow \mathrm{R}yx)$

(3) $\forall x\forall y\forall z((\mathrm{R}xy \,\&\, \mathrm{R}yz) \rightarrow \mathrm{R}xz)$

interactions with the referents of those terms and this the key to Frege's solution to the problem of our knowledge of abstract objects. Our proposed definitions of the operators Dir(x) and Num(F) however, fail to specify determinate truth conditions for *all* sentences in which those terms can occur and hence the context principle cannot be applied - the definitions do not uniquely fix the references of those terms.

For this reason, Frege rejects the proposed contextual definition of the direction operator and by analogy, the proposal to take Hume's Principle as a contextual definition of the cardinality operator. In §68 then, Frege adopts an alternative type of definition. He now shows how we can give *explicit* definitions of the direction and cardinality operators, using the concept of the *extension* of a concept.

Think of the extension of a concept F as the set of all and only those objects which have the property F. If we abbreviate the phrase 'the extension of the concept F' as $\mathrm{Ext}_x(Fx)$ - read this as 'the set of all objects x such that Fx' - then Frege's proposed definitions of the direction and cardinality operators are as follows:

(Def. Dir) $\quad \mathrm{Dir}(a) =_{df} \mathrm{Ext}_x(x \mathbin{//} a)$

(Def. Num) $\quad \mathrm{Num}(F) =_{df} \mathrm{Ext}_x(x \approx F)$

These are explicit definitions of the operators involved, having the form of identities. According to the first, the direction of a is the extension of the concept 'is parallel to a'; alternatively, the direction of a is the set of all lines parallel to a. According to the second, the number belonging to a concept F is the extension of the concept 'is equinumerous with F'; alternatively, the number belonging to a concept F is the set of all concepts which are equinumerous with F. Take some arbitrary concept F and suppose it applies to just three things; a, b and c. The extension of the concept F is then just the set $\{a, b, c\}$. The number belonging to the concept F is, according to Frege's definition, the set of all concepts {F, G, H, ...} which can be put in one-one correspondence with F. This is Frege's number three - the set of all concepts which have exactly three objects falling under them.

§§70-83 contain the details of Frege's derivation of the basic laws of arithmetic from his definitions of arithmetical concepts. Frege first shows how the notion of a one-one correlation can be defined in purely logical terms, using only first and second-order quantifiers and identity. In §§70-71 he says that a relation R provides a one-one correlation of the Fs with the Gs just in case:

(A) Every F bears the relation R to exactly one G.

(B) Every G is related by R to exactly one F.

In modern notation, we can express (A) and (B) as:

(Def. $F\ 1\text{-}1_R\ G$) $\quad \forall x(Fx \rightarrow \exists_1 y(Gy \ \&\ Rxy)) \ \&\ \forall x(Gx \rightarrow \exists_1 y(Fy \ \&\ Ryx))$

where the quantifier $\exists_1$ can be defined in the usual way:

(Def. $\exists_1$) $\quad \exists_1 y(\Phi y) =_{df} \exists y(\Phi y \ \&\ \forall z(\Phi z \rightarrow z = y))$

In §73, Frege sketches a proof of Hume's Principle from his explicit definition of the cardinality operator (Def. Num above). This is the first and only time he makes use of that definition, all the proofs that follow depend only on Hume's Principle. I will examine Frege's attempted proof of Hume's Principle in the next section.

Frege now shows how to define the terms *zero*, *successor* and *natural number*. The number zero is defined to be the number belonging to the concept 'is not identical with itself' [§74]. That is:

(Def. 0) $0 =_{df} \mathrm{Num}_x(x \neq x)$

For the definition of zero, Frege could have chosen the number belonging to any concept which has no instances, for example, the number belonging to the concept 'is a unicorn'. The advantage of choosing the concept 'is not identical with itself' of course, is that it is a *logical* truth that this concept has no instances.

The successor relation is defined in §76. Frege there defines 'n is a successor of m' to mean 'for some concept F, n is the number of Fs and there is an object x which is F and m is the number of things which are F, but distinct from x'. In symbols, we have:

(Def. S) $Snm =_{df} \exists F \exists x(Fx \;\&\; n{=}\mathrm{Num}_y(Fy) \;\&\; m{=}\mathrm{Num}_z(Fz \;\&\; z \neq x))$

From this definition and Hume's Principle, we can prove that successor is a one-one relation; the fourth Dedekind-Peano axiom, stating that no distinct natural numbers have the same successor is an immediate consequence. From the definitions of successor and zero, it is easy to prove that for all x, zero is not the successor of x; the second Dedekind-Peano axiom, that zero is not the successor of any natural number, is an obvious consequence of this theorem.

Frege's definition of *natural number* is a little more complex. Frege first defines the relation 'n follows m in the series of natural numbers'. Following Dummett[17], I will abbreviate this as $n > m$. To define this relation, Frege makes use of the notion of the *ancestral* of a relation, which he had defined in the *Begriffsschrift* [§9]. The definition of

[17] See [Dummett 1991, pp. 119-124]

this notion given in §79 of *Grundlagen* can be unpacked into two parts. Firstly, let us say that a property F is *hereditary with respect to a relation* R iff whenever an object x has the property F and x bears the relation R to y, then y also has the property F. That is:

(Def. HER) $\mathrm{HER}(F, R) =_{df} \forall x \forall y((Fx \;\&\; Rxy) \rightarrow Fy)$

Secondly, let us say that a property F is *inherited by every object after* a *in the R-series* iff whenever a bears the relation R to an object x, x has the property F:

(Def. IN) $\mathrm{IN}(F, a, R) =_{df} \forall x(Rax \rightarrow Fx)$

Frege's definition of 'a is the ancestral of b with respect to R',[18] can then be expressed as 'b has every property which is hereditary with respect to R and inherited by every object after a in the R-series':

(Def. ANC) $\mathrm{ANC}(a, b, R) =_{df} \forall F((\mathrm{HER}(F, R) \;\&\; \mathrm{IN}(F, a, R)) \rightarrow Fb)$

Frege's procedure is then to define $n > m$ ('n follows m in the series of natural numbers') to mean 'm is an ancestral of n with respect to the *converse* of the successor relation' [§81]. That is:

(Def. >) $n > m =_{df} \mathrm{ANC}(m, n, P)$

Where P is the *predecessor* relation; $Pxy =_{df} Syx$. Frege now defines the relation 'n belongs to the series of natural numbers beginning with m'. Again following Dummett, I write this as $n \geq m$. This is defined in §81 to mean 'either $n > m$ or $n = m$':

(Def. ≥) $n \geq m =_{df} n > m \vee n = m$

Frege's definition of *natural number* can now be stated quite simply; n is a natural number iff n belongs to the series of natural numbers beginning with zero [§83].[19] That is:

[18] Frege's terminology for this is 'b follows in the R-series after a'.
[19] Frege uses the expression *finite number* in place of *natural number*.

(Def. NAT) $\text{NAT}(n) =_{df} n \geq 0$

The effect of this definition is that n is a natural number just in case n is either zero or can be reached by starting from zero and going from one number to its successor. The first Dedekind-Peano axiom, stating that zero is a natural number is clearly an immediate consequence. More surprisingly, the principle of mathematical induction, the fifth Dedekind-Peano axiom, can be quite easily proved from this definition. This represents something of an achievement on Frege's part. He has shown, *pace* Poincaré, that the principle of mathematical induction can be seen as simply a consequence of the very definition of *natural number* and is therefore *analytic* - a second-order logical truth.

This leaves only the third axiom, which states that every natural number has a successor, which is also a natural number. It is easy to prove that the successor of any natural number, if there is one, is also a natural number, using Frege's definitions of successor and natural number. The claim that every natural number *has* a successor is not so easy to prove. This was always going to be the biggest hurdle for Frege's programme; the task of showing that *pure logic* can tell us that there are *infinitely* many objects.

How does he do it? Firstly, he has shown that the number *zero* exists; this is the number belonging to the concept 'is not identical with itself'. Now consider the concept, 'is a natural number less than or equal to zero'. Exactly one thing falls under this concept, (namely the number zero) and so Frege defines the number *one* to be the number belonging to this concept. Hence the number one exists. Then we take the concept 'is a natural number less than or equal to one'. Exactly two things fall under this concept (the numbers one and zero) and so Frege defines the number *two* as the number belonging to this concept. Hence the number two exists. Obviously, we can iterate this process to infinity.

Frege sketches his proof that every natural number has a successor in §§82-83 of *Grundlagen*. Frege has to provide, for every number n, a concept such that the number of things falling under it is the successor of n. The reasoning given above suggests the concept 'natural number $\leq n$', where $\leq$ is just the converse of the relation $\geq$ defined above; $a \leq b =_{df} b \geq a$. There is one natural number ≤ 0; and one is the successor of 0. Likewise, there are two natural numbers ≤ 1 and two is the successor of 1. There are three numbers ≤ 2, and three is the successor of two. Frege shows how we can prove by mathematical induction, the general theorem, that the number belonging to the concept 'natural number $\leq n$' is the successor of n. From this, the desired theorem that every natural number has a successor follows at once by existential generalisation.

It seems then, as though Frege has established at least the first two components of his programme. He has defined the basic concepts of arithmetic in purely logical vocabulary and derived the basic laws of arithmetic from those definitions. Unfortunately, as we now know, things are not so simple. But where did Frege go wrong?

3. Why Frege's Programme Failed

Recall that Frege's final explicit definition of the cardinality operator made use of the notion of the extension of a concept. He defined the number belonging to a concept F as the extension of the concept 'is equinumerous with F'. That is:

(Def. Num) $\text{Num(F)} =_{df} \text{Ext}_x(x \approx \text{F})$

Intuitively, the extension of a concept is the set of all objects which fall under the concept. So according to Frege's definition, the number belonging to the concept F is the set of all concepts equinumerous with F. Unfortunately for Frege, the introduction of extensions was to lead to disaster. In *Grundlagen*, Frege does not put forward any specific theory of extensions, contenting himself with the remark "I assume that is known what the extension of a concept is." [§69, footnote 2]. This represents a significant lacuna in the argument of *Grundlagen*. Like numbers, extensions are themselves abstract objects, so in order to complete Frege's programme, we need to provide a logical justification for the introduction of the *extension operator* $\mathrm{Ext}_x(Fx)$. In the light of the context principle, this would involve laying down determinate truth conditions for all sentences involving such terms. Since extensions are objects, we need, in particular, to state the truth conditions for identity statements involving extensions.

Frege's theory of extensions was presented formally in *Grundgesetze Der Arithmetik* [Frege 1893-1903]. There, Frege introduces the extension operator in a manner analogous to the proposed *contextual* definitions of the direction and numerical operators which he had rejected in the *Grundlagen*; that is, he defines the operator by stating an identity condition for extensions in terms of an equivalence relation. Axiom V of *Grundgesetze* states, in effect, that the extension of the concept F is identical with the extension of the concept G if and only every F is a G and every G is an F - in other words, if and only if, F and G are *co-extensive*.[20] In symbols, we have:

(Axiom V) $\forall\Psi\forall\Phi\,(\,\mathrm{Ext}(\Psi) = \mathrm{Ext}(\Phi) \leftrightarrow \forall x(\Psi x \leftrightarrow \Phi x)\,)$

[20] I have simplified somewhat here. In fact Frege's Axiom V is more general than this. It applies to the extensions ('range-of-values') of arbitrary functions, and not just to the special functions from objects to truth-values which represent concepts.

In a letter to Frege, written just before the second volume of *Grundgesetze* was due to be published, Bertrand Russell showed that Axiom V was inconsistent. In the context of Frege's system, Russell's paradox arises in the following way.

In a second-order logic with identity and $\mathrm{Ext}(\Phi)$ as a primitive function symbol, we can prove the following 'comprehension principle' for extensions:

(Comprehension) $\forall\Phi\exists x(\,x = \mathrm{Ext}(\Phi)\,)$

This is a second-order logical truth, stating that every concept has an extension. Really all this amounts to is that the function referred to by $\mathrm{Ext}(\Phi)$ is defined (has a value) for every argument. This becomes a logical truth because it is a requirement on every interpretation of a function symbol, that the function is defined for every object in the domain of discourse.[21]

Consider the concept 'x is the extension of some concept which x itself falls under'. We define:

$$Sx =_{df} \exists\Phi(\,x = \mathrm{Ext}(\Phi)\ \&\ \Phi x\,)$$

The extension of this concept will be the set of all things which are the extensions of some concept they themselves fall under. An example of an object which has the property S is the extension of the concept 'is an extension'. Call the extension of this concept s. Then obviously s is the extension of some concept, namely S - the concept 'is an extension'. Equally obviously, s itself falls under this concept, since s itself is an extension.

[21] The proof of this comprehension principle is very simple. Let F be an arbitrary concept. Then, by the laws of identity, we have Ext(F) = Ext(F). By existential generalisation, it follows that $\exists x(\,x = \mathrm{Ext}(F)\,)$. Since F was arbitrary, we can deduce: $\forall\Phi\exists x(\,x = \mathrm{Ext}(\Phi)\,)$.

Now consider the concept which holds of just those things for which Sx does not hold. We define:

$$\mathrm{R}x =_{df} \sim\mathrm{S}x = \sim\exists\Phi(\, x = \mathrm{Ext}(\Phi)\ \&\ \Phi x\,)$$

Equivalently, we can define Rx as:

(Def. R) $\mathrm{R}x =_{df} \forall\Phi(\, x = \mathrm{Ext}(\Phi) \rightarrow \sim\Phi x\,)$

This is the Russell property. The extension of R will be the set of all things which are not the extensions of any concepts they themselves fall under. Anything which is not an extension will have the property R. Some extensions will also have it. An example would be the extension of the concept 'is a man', since this object is of course not itself a man.

We can now show how Axiom V leads to contradiction. Consider the concept R, defined as above. By the comprehension principle, this concept has an extension, which we will denote by the symbol r:

(1) $r = \mathrm{Ext}(\mathrm{R})$

We now ask whether r has the Russell property or not. Suppose it does. That is, suppose:

(2) $\mathrm{R}r$

Then, by the definition of R, we have:

(3) $\forall\Phi(\, r = \mathrm{Ext}(\Phi) \rightarrow \sim\Phi r\,)$

Since this holds for every concept Φ, it holds in particular for R:

(4) $r = \mathrm{Ext}(\mathrm{R}) \rightarrow \sim\mathrm{R}r$

Applying *modus ponens* to (1) and (4), we have:

(5) $\sim\mathrm{R}r$

The assumption that r has the Russell property, led to the conclusion that r does not have that property. Hence:

(6) $\mathrm{R}r \rightarrow \sim\mathrm{R}r$

So far, we have not made use of Axiom V at all. We now use it to show that, conversely, if r does not have the Russell property, then r does have that property. We suppose:

(7) $\sim\mathrm{R}r$

Applying the definition of R, we get:

(8) $\exists\Phi(\, r = \mathrm{Ext}(\Phi)\ \&\ \Phi r\,)$

This says that r is the extension of some concept which r itself falls under. Call this concept G. Then we have:

(9) $r = \mathrm{Ext}(\mathrm{G})$

(10) $\mathrm{G}r$

But, by our initial assumption (1) $r = \mathrm{Ext}(\mathrm{R})$. So from (1) and (9) we have:

(11) $\mathrm{Ext}(\mathrm{R}) = \mathrm{Ext}(\mathrm{G})$

By Axiom V, (11) is equivalent to:

(12) $\forall x(\, \mathrm{R}x \leftrightarrow \mathrm{G}x\,)$

In particular, we have:

(13) $\mathrm{R}r \leftrightarrow \mathrm{G}r$

From (10) and (13), we obtain the desired result:

(14) $\mathrm{R}r$

So from the assumption that r does not have the Russell property, it follows that it does have it:

(15) $\sim\mathrm{R}r \rightarrow \mathrm{R}r$

Combining (15) and (6), we deduce:

(16) $\mathrm{R}r \leftrightarrow \sim\mathrm{R}r$

This is a contradiction. So Axiom V is inconsistent.

How does this effect Frege's derivation of the basic laws of arithmetic? Recall that Frege attempted to use his explicit definition of the cardinality operator to prove Hume's Principle:

(Hume's Principle) $\forall\Psi\forall\Phi(\,\mathrm{Num}(\Psi) = \mathrm{Num}(\Phi) \leftrightarrow (\Psi \approx \Phi)\,)$

where $\Psi \approx \Phi$ means 'there is a one-one correlation between the Ψs and the Φs'. In fact, as already remarked, this is the only use Frege ever makes of the definition of the cardinality operator in terms of extensions. The proofs of the Dedekind-Peano axioms depend only on Hume's Principle.

Let us see how Frege's proof of Hume's Principle from the definition of the cardinality operator is meant to work. We need to prove two things, for arbitrary concepts F and G:

(A) $\mathrm{F} \approx \mathrm{G} \rightarrow \mathrm{Num(F)} = \mathrm{Num(G)}$

and

(B) $\mathrm{Num(F)} = \mathrm{Num(G)} \rightarrow \mathrm{F} \approx \mathrm{G}$

Let us consider how we would prove (A). We start by assuming:

(1) $\mathrm{F} \approx \mathrm{G}$

and we need to prove that Num(F) = Num(G). Given that $\approx$ is an equivalence relation, it is straightforward to show that (1) entails:

(2) $\forall x(x \approx \mathrm{F} \leftrightarrow x \approx \mathrm{G})$

Frege then claims that (2) entails:

(3) $\mathrm{Ext}_x(x \approx \mathrm{F}) = \mathrm{Ext}_x(x \approx \mathrm{G})$

By the definition of the cardinality operator, we have $\mathrm{Num(F)} = \mathrm{Ext}_x(x \approx \mathrm{F})$ and $\mathrm{Num(G)} = \mathrm{Ext}_x(x \approx \mathrm{G})$. Therefore, (3) entails:

(4) $\mathrm{Num(F)} = \mathrm{Num(G)}$

as required. The crucial step here is that from (2) to (3). This is the point at which Frege made an implicit appeal to Axiom V. Substituting the concepts $x \approx \mathrm{F}$ for Ψ and $x \approx \mathrm{G}$ for Φ in that Axiom V, we obtain:

(5) $\mathrm{Ext}_x(x \approx \mathrm{F}) = \mathrm{Ext}_x(x \approx \mathrm{G}) \leftrightarrow \forall x(\,x \approx \mathrm{F} \leftrightarrow x \approx \mathrm{G}\,)$

and given (5), (2) obviously entails (3). But since Axiom V is inconsistent, Frege's derivation of Hume's Principle from his definition of the cardinality operator is unsound. Frege's strategy for showing arithmetic to be analytic was in ruins. A principle indispensable to his derivation of the basic laws of arithmetic had been shown to be inconsistent.

Nonetheless, Frege's technical achievement should not be underrated. The proofs of the five Dedekind-Peano axioms from Hume's Principle are perfectly sound. The fact that Hume's Principle, given appropriate definitions, entails the Dedekind-Peano axioms for arithmetic has become known as *Frege's Theorem*.[22] The second-order Dedekind-Peano axioms are categorical - all the models of the theory are isomorphic to each other - and every truth of arithmetic is a logical consequence of those axioms.[23] Frege's Theorem shows then, that there is a sense in which the entire body of number theory can be founded a single, self-evident principle which states an identity condition for numbers.

---

[22] See [Demopoulos 1995, *passim*].

[23] Of course, since second-order logic is incomplete, it does not follow that every truth of arithmetic is *provable* from those axioms.

Crispin Wright has used this fact to argue that we can salvage Frege's logicism from the wreckage of the contradiction.[24] If Hume's Principle were a logical truth, Frege's logicism would indeed be unaffected by the inconsistency in the theory of extensions. But Hume's Principle is not a logical truth in the standard sense; it is not true under every allowable interpretation of the function symbol Num(F).

To some extent of course, the question of whether Hume's Principle is a logical law or not is a terminological one. Leibniz's law is not a logical truth in this sense either; it is not true under every interpretation of the relation symbol =. Nonetheless, it is standard practice to include in systems of logic the special symbol = taken to be governed by a special 'logical axiom', namely Leibniz's law. This fits quite well with Frege's characterisation of the laws of logic as involving only terms of universal application; the concept of identity is so general that it applies to objects of any kind whatsoever. It could be argued that we should think of the cardinality operator in the same way. Since objects of any kind whatsoever can be numbered, it follows that the cardinality operator is also universally applicable. Why then should we not introduce the special function symbol Num(F) into our second-order logic, and include Hume's Principle as a logical axiom governing its use? I think we could certainly do this and that it would not be completely unreasonable to call the resulting system a *logic*. If so, we might say that under this wider application of the term *logic*, Frege was correct; arithmetic is reducible to logic, since it is reducible to Hume's Principle.[25] But the important question is, what would be the epistemological significance of this move?

---

[24] In *Frege's Conception of Numbers As Objects* [Wright 1983].
[25] Dummett makes a similar point in *Frege: Philosophy of Mathematics*: "The style of objection to logicism now exceedingly frequent is therefore quite beside the point: the objection for instance, that set theory is not part of logic....By Frege's criterion of universal applicability, the notion of cardinal number is *already* a logical one, and does not need the definition in terms of classes to make it so...The definition in terms of classes is *not* needed to show arithmetic to be a branch of logic...Had Frege been concerned only with number theory...and had he been able to solve the Julius Caesar problem for numbers...then it would not have impaired his logicist programme to take the numerical operator as primitive" [Dummett 1991, p. 224-5].

Crispin Wright does not take this line. Instead adopts the proposal, rejected by Frege, that we take Hume's Principle as constituting a definition, albeit a contextual one, of the cardinality operator. He argues that definitions which introduce terms for abstract objects by stating a criterion of identity for them in terms of an equivalence relation on objects of some other kind ('definition by abstraction' as it is often called) is legitimate and that the Julius Caesar problem can be solved. If so, the way is open for a logical justification of arithmetic by means of the context principle. But is this kind of definition really legitimate? Michael Dummett has argued that it cannot be. For, if it were, then the method Frege uses to introduce the extension operator would also be legitimate, since it has exactly the same form. But we know that it in this case the technique was not legitimate, since it led to a contradiction.[26]

This point raises the question of whether Hume's Principle itself is consistent. Since Hume's Principle has an analogous logical form to Axiom V, it might be possible to construct an analogous proof of an inconsistency. Wright considers this, but argues that the proof will not go through. Again, as before, we have as a second-order logical truth, a comprehension principle for numbers:

$$\forall\Phi\exists x(\, x = \mathrm{Num}(\Phi)\,)$$

which states that every concept has a number. We could again define a concept R analogously as:

$$Rx =_{df} \forall\Phi(x = \mathrm{Num}(\Phi) \rightarrow \sim\Phi x)$$

---

[26] See [Dummett 1991, pp. 187-9].

An object has the property R if it is not the number of any concept it falls under. Obviously, the first half of the proof given above of the inconsistency of Axiom V, will go through without change. By comprehension for numbers, the concept R has a number, call it r. So we have:

(1) $r = \text{Num}(R)$

As before, we will be able to derive $Rr \rightarrow \sim Rr$. However, the next half of the proof fails. From the assumption that $\sim Rr$, we can prove that there must be *some* object which has the property R, but of course this does nor entail that r has the property R. So the derivation of Rr from the assumption that $\sim Rr$ is blocked.[27] Of course, this argument does not show that Hume's Principle must be consistent, it only shows that one route to a contradiction is blocked. We now know however, that Hume's principle *can* in fact be proved consistent.[28]

Nonetheless, Dummett's objection to Wright still stands. The inconsistency of Axiom V shows that proposed method for contextually defining terms for abstract objects is not, *in general*, a legitimate technique. Wright cannot get himself off the hook by suggesting that the method is legitimate, provided only that the definition is consistent. For there are consistent principles of this form which are inconsistent with each other. George Boolos gives an example in 'The Standard of Equality of Numbers' [Boolos 1995b, pp. 250-1; see also Hazen 1985]. Say that the concepts F and G 'differ evenly' if the number of objects falling under F, but not under G, or the number of objects falling under G, but not F, is a finite even number. This relation between concepts is an equivalence relation and can be defined in purely logical (second-order) vocabulary. Now introduce the operator 'the parity of F' by means of the *Parity Principle*; the parity of F = the parity of G iff F and G

[27] See [Wright 1983, pp. 155-156].
[28] See for example [Boolos 1995a, pp. 216-217].

differ evenly. The Parity Principle is consistent, but Boolos shows it is false in any infinite domain. Hence it is inconsistent with Hume's Principle, which entails the existence of infinitely many objects, as Frege proved. Both principles then are consistent and of the required form, but they cannot both be true, since they are inconsistent with each other. The prospects appear bleak then, for this attempt to salvage Frege's programme. In the remaining sections, I argue that indeed, *any* attempt to provide mathematics with foundations is epistemologically misguided.

4. On Certainty

We have seen then that Frege's programme failed for mathematical reasons. This is a feature it shares with one of the other great early twentieth century foundational programmes; the formalist programme of Hilbert, which was fairly conclusively refuted by Gödel's second incompleteness theorem.[29] Since these programmes failed for quite specific technical reasons, their failure may seem to leave open that possibility that some foundational programme might succeed where they have failed. I would like now to discuss some of the epistemological presuppositions underlying the doctrine of foundationalism; to see if there are any general objections to the idea that arithmetic, or any other branch of mathematics, has or needs to be provided with foundations.

As I am using it here, foundationalism with respect to a given subject matter is the doctrine that there exists a set of epistemologically privileged first principles which serve to

[29] The other two foundational programmes are of course, the development of Frege's logicist programme along positivist lines initiated by Russell and Whitehead [Russell and Whitehead 1910-13; see also Carnap 1931] and the Intuitionist programme of Brower and Heyting [Brower 1913;1949, Heyting 1931]. I will not discuss these programmes in any detail here; the reader may decide how the arguments developed here apply to these cases.

justify every other truth of that subject matter. Foundationalists differ as to the nature and status of the first principles and in the kind of justification they provide. For Frege, the foundational propositions of arithmetic were the truths of logic and the justification deductive - he aimed to show that every truth of arithmetic could be deduced from logical laws, by means of definitions. For Hilbert, the foundational propositions were *finitary statements*; decidable propositions which do not involve reference to infinite totalities. The rest of mathematics, consisting of *ideal statements*, Hilbert supposed to be strictly meaningless. Finitary arithmetic, in Hilbert's programme, provides the sure and certain foundation. The ideal statements of arithmetic are justified by showing that if we add them to the foundation, no new finitary statement can be deduced. This is to say that the classical arithmetic is a *conservative extension* of finitary arithmetic. Even though adding the ideal statements does not allow us to prove anything new, they nonetheless have an instrumental justification - they provide us with an extremely useful means of proving facts about the finitary subject matter of mathematics; facts we could prove without them, though not as concisely or as elegantly. Furthermore, since a proof of conservativeness is in this case equivalent to a proof of consistency, we can be sure that the introduction of the ideal statements will never lead us into contradiction.[30]

We have two different kinds of foundational proposition here, logical laws in Frege's case, finitary statements in Hilbert's. In both cases, it is important that the foundational propositions are taken to have an epistemologically privileged status, in the sense that they are truths we can be especially certain of. Foundationalism is, in this sense, a search for certainty. But one may question whether such a search is justifiable.

[30] See [Hilbert 1926, Giaquinto 1983].

Frege had aimed to found arithmetic on the certainty of logic, by showing that our knowledge of arithmetic is derived from our knowledge of the fundamental laws of logic. Nowhere in *Grundlagen* does Frege address the question of how we know the laws of logic to be true. It has been argued that he felt no need to, believing the answer to this question to have been already settled by Kant.[31] Consider Frege's classification of truths into *a priori* and *a posteriori*, analytic and synthetic. What is the basis of Frege's classification? Why should we call 'analytic' those propositions deducible from logical laws and 'synthetic' those deducible only from laws of some special science? Why call either of these '*a priori*'?

Although there are obvious differences here between Frege's distinctions and the corresponding distinctions in Kant, Frege says that he does not "mean to assign a new sense to these terms but only to state accurately what earlier writers, Kant in particular, have meant by them" [§3]. For Kant, the distinction between the *a priori* and the *a posteriori* is epistemic; *a priori* propositions can be known independently of experience, *a posteriori* truths cannot. But Kant, unlike Frege, drew the analytic/synthetic distinction in terms of the *content* of a proposition, rather than in terms of its justification; an analytic proposition is one where the concept of the subject contains the concept of the predicate (Kant's example is 'all bodies are extended'), a synthetic proposition is one in which although the predicate may apply to the subject, the concept of the subject does not contain the concept of the predicate ('all bodies are heavy').[32] Frege saw that since Kant was wrong in supposing that every proposition is of subject-predicate form, his distinctions failed to be exhaustive;

[31] See [Kitcher 1979].
[32] See [Kant 1881: Introduction, §§ iv-v].

propositions not of that form will not be classified under Kant's scheme. Hence a redefinition of the Kantian distinctions is required.

Frege's idea was to define both of the Kantian distinctions in epistemic terms, by classifying all truths in terms of the their justifications. This justification remember, is always a deductive proof from unprovable first principles, so the question becomes one of classifying the source of our knowledge of the first principles. Kant held that in general, there are just three sources of knowledge. Every proposition is either (1) synthetic *a posteriori*, (2) analytic *a priori* or (3) synthetic *a priori*. Our knowledge of truths of type (1) comes from sensible intuition, or perception. *A priori* knowledge on the other hand, is independent of experience. Our knowledge of truths of type (2) comes from a kind of conceptual analysis; we have a means of recognising when one concept is 'contained' in another, which is independent of experience, but not further elucidated by Kant. Our knowledge of truths of type (3) come from what Kant called a *pure intuition.*

Frege accepted this three-fold classification of the sources of knowledge and used it as the basis for his redefinition of the Kantian distinctions. In Frege's scheme, *a posteriori* truths are ones which can only be deduced from a primitive truth which ascribes a property to a particular object and such facts are known by observation or perception. *A priori* truths are ones which can be deduced from general laws alone. Our knowledge of these laws is not based on experience or perception. Some of these general laws apply only to things which can exist in space or time and our knowledge of them is grounded in a pure intuition (spatial or temporal). These are the synthetic *a priori* truths. Other general laws are of universal applicability, they apply to any object whatsoever, even objects we cannot perceive or intuit. Hence our knowledge of them cannot be based on intuition or perception.

These are the analytic *a priori* truths. Where Kant had spoken of an ability to recognise relations of containment between concepts, Frege widens the idea so as to include statements not of subject-predicate form. He speaks instead of the 'logical source of knowledge' or the 'logical faculty'.[33]

To summarise; for Frege, primitive *a posteriori* truths are known by observation or perception, primitive synthetic *a priori* truths, like the axioms of geometry are known by pure intuition and primitive analytic truths, like the basic laws of logic, are known by means of the operation of the logical faculty; a process which provides us with knowledge that is independent of both perception and intuition.[34] All other truths are known by deducing them from primitive truths known in one of these three ways.

Just what is this 'logical faculty' which provides us with knowledge of the laws of logic? Frege provides no answer to this question. It is not discussed at all in the *Grundlagen*. In the *Grungesetze* he remarks only that "[t]he question why and with what right we acknowledge a law of logic to be true, logic can answer only by reducing it to another law of logic. Where that is not possible logic can give no answer." [Frege 1893-1903]. Frege's silence on this issue may be traced to his view that the question how we come to know the laws of logic is something for *psychology* to answer; it is not a question for mathematics or logic. There must be *some* psychological process by means of which we come to know the laws of logic, but Frege does not see it as his task to explain how this process works.

[33] See [Kitcher 1979; Dummett 1991 p. 44].
[34] Hence, as Frege remarks in §4 of *Grundlagen*, the mathematical project of eliminating appeals to intuition (spatial or temporal) from arithmetic and the philosophical problem of showing that arithmetic is analytic come to the same thing.

Nonetheless, Frege accepted the Kantian doctrine that whatever the nature of this process, it can provide us with knowledge that is independent of experience and absolutely certain. It is this unexpressed presupposition of Frege's thought that provides the central epistemological motivation to his programme. The contradiction in Frege's system was so devastating to him because it casts doubt on the presupposition.

Let us keep the phrase 'the logical faculty' as a shorthand way of referring to the source, whatever it may be, of our logical knowledge. What can we say about it? In fact our beliefs concerning logical truths have two main sources. Some truths of logic we accept because they are 'intuitively obvious', others we believe only because they form part of a logical theory. Even before we begin to study formal logic, we have certain pre-theoretic intuitions about the validity of certain inferences. It is these that we often appeal to in teaching logic to students and they are sometimes appealed to in criticisms of logical theories. But these intuitions can hardly be said to give us certain knowledge. It is well known to psychologists that our logical intuitions can often lead us into error. People often make mistakes in simple deductive reasoning. Frege's logical faculty may have convinced him of the truth of his ill-fated Axiom V; the inconsistency of that axiom is testament to the fact that what seems intuitively logically self-evident can often be false. Whatever the psychological source of our pre-theoretic judgements concerning logical truths, Frege and Kant were surely wrong in supposing that it provides an absolutely reliable belief forming mechanism.

Furthermore, our logical intuitions are not only sometimes faulty, they often simply fail to give a verdict. Where intuition fails, theory has to take over. In the words of logicians R. K. Meyer and J.K. Slaney:

> Consider "If if if snow is white then grass is green then grass is green then snow is white." If you, dear reader, have any firm intuitive sense of whether this is valid, then your intuitions carry you further into left field than do ours. Such sentences get labelled "logically valid" or "logically invalid" only on the basis of some theory, not on whether they commend themselves individually to their beholders.
>
> [Meyer and Slaney 1989, p. 256]

This is the second source of our beliefs concerning the laws of logic. There are many logical principles which we accept or reject only the basis of some systematic theory of deductive inference. Of course, Frege is well aware of this. For him, our logical faculty provides us with knowledge of the unprovable basic axioms of logic, which are simple and self-evident. More complex logical truths, like the one above, are known by deducing them from such axioms. But the relationship between the axioms of the logical theories we accept and our pre-theoretic judgements is not so simple.

Intuitions, even concerning fundamental axioms, can be modified in the light of improvements in logical theories. In Aristotle's logic for example, the inference from a universal statement (All men are mortal) to the corresponding existential (Some men are mortal) is valid. The inference seems intuitively valid to many people – the concept of a vacuously true universal statement is somewhat counter-intuitive. Nonetheless, the inference is rejected in modern logic. A logical theory may account for some of our pre-theoretic judgements or intuitions and be in conflict with others. Where a logical theory conflicts with intuition, we may come to abandon the intuition and accept what seemed to be a counter-intuitive consequence of the theory. We are justified in doing so to the extent that the theory gives a plausible and *systematic* account of the laws of logic. The classical definition of valid inference, for example, enables us to give a uniform account of the validity of a wide range of particular inferences. It does have some counter-intuitive consequences (that a contradiction entails everything; that everything entails a logical truth and so on) but we may come to accept these because they are consequences of a systematic theory we have good grounds for accepting. In other cases, the counter-intuitive consequences of a logical theory may lead us to reject the theory and adopt another. Such is one motivation for relevance logic, for example. But the fact that a theory conflicts with some our intuitions is not, by itself, conclusive evidence that the theory is incorrect. In fact, there a few intuitions so indubitable that some logician or other has not at some time denied them. Logicians there are who have denied even the law of non-contradiction, that no statement is both true and false.[35]

It is not that we have a sure and certain grip on certain fundamental logical axioms and all the rest of our knowledge is deduced from these. Rather, we accept axioms to the extent that they give us a systematic account of deductive inference; an account which may justify many of our pre-theoretic judgements, but also show us that some of them are false. There is a complex interplay between theoretical and pre-theoretical judgements. Intuitions are not sacrosanct and neither are logical theories. Both are revisable; intuitions are revisable in the light of theoretical principles and those theoretical principles themselves are revisable, not only in the light of our uncertain intuitions, but also the grounds that an alternative account provides a more systematic or powerful theory of deductive inference.[36]

The sources of our logical knowledge are therefore not very different from the sources of our knowledge of other kinds. They do not provide us with an especially secure

[35] See for example [Priest 1987].

[36] Some have even argued that logical theories can even be revised in the light of *empirical* evidence. [Quine 1951, Putnam 1971]. This argument is independent of the one given here – I examine it in more detail in chapter four. If it is correct, then Frege and Kant were wrong not only in supposing that our logical knowledge is of an absolutely certain kind, they were also wrong in believing it to be independent of experience.

kind of knowledge, but only with beliefs which are both fallible and revisable. The logical faculty whatever its nature, is an imperfect source of knowledge, it can lead us into error as well as into truth. Logic then, cannot provide an absolutely certain foundation on which to base a mathematical theory like arithmetic.[37]

Hilbert, by contrast, wanted to found arithmetic on the certainty of finitary statements. The programme was to show that classical arithmetic is a conservative extension of finitary arithmetic. This is equivalent to showing, using only finitary reasoning, that classical arithmetic is consistent. If this could be done, then we would have a guarantee that the introduction of ideal (non-finitary) statements into arithmetic can never lead to contradiction.

As already remarked, Gödel's second incompleteness theorem revealed that Hilbert's programme was unachievable. According to that theorem, classical arithmetic, if it is consistent, cannot prove its own consistency.[38] Hence finitary arithmetic, being a proper part of classical arithmetic, cannot provide such a consistency proof either.

The point can be generalised. Suppose you were worried about the certainty of arithmetic. You want to be able to prove that arithmetic is at least consistent. To prove that arithmetic is consistent you will need a theory T in which to construct the proof. Gödel's theorem tells us that the theory T will have to be stronger than arithmetic itself; T cannot be a proper subset of arithmetic. But if you were worried about arithmetic, you ought also to be worried about any mathematical theory that goes beyond arithmetic. Hence you ought

[37] Frege himself seems to have come to recognise this. In a late paper, published in his *Posthumous Writings*, he reiterates the validity of the Kantian three-fold classification of sources of knowledge; perception, pure spatial or temporal intuition and the logical source of knowledge. However, he argues that contrary to his previous belief, the logical source of knowledge is imperfect and prone to error. He continues to believe that our knowledge of arithmetic is not based on perception. The only alternative, if we want to retain a privileged status for arithmetical knowledge is to base it on pure intuition, and so Frege sets about investigating the possibility of reducing arithmetic, not to logic, but to *geometry*. [Frege 1979, see also Kitcher 1979].

[38] See [Gödel 1931].

also to be worried about the consistency of T. The consistency of arithmetic can in fact be proved, but it can *only* be proved by means of some fairly high-powered transfinite set-theory.[39] Such a proof is worthless if absolute certainty is your aim; if arithmetic is open to doubt, transfinite set theory cannot be any better off. Hence, the search for absolute certainty regarding arithmetic is obtainable.

We can now see more clearly why there would be little epistemological gain for a foundationalist in adopting idea mentioned in section three for salvaging Frege's programme. We could take the cardinality operator as primitive and governed by Hume's Principle as an axiom. We could even, with some justification, call the resulting system a *logic*. But this would not show that our knowledge of arithmetic is grounded in anything we can be especially certain of. Whether we call Hume's Principle a definition or not, we only know that it is *consistent* by means of a proof which appeals to principles of higher set-theory, principles which cannot be more certain than arithmetic itself. Neither logic, nor any privileged part of mathematics itself can provide us with the certainty yearned for by foundationalism. The quest for absolute certainty in mathematics, as in the rest of science, should be abandoned.

5. On Proper Justification

I have not yet discussed the second epistemological motivation for Frege's programme. Recall that Frege was not only concerned to establish the *certainty* of arithmetic, he also wanted to exhibit the true grounds on which it should be accepted; its *proper justification*. In particular, he wanted to show that the proper justification of

[39] See [Gentzen 1943].

arithmetic was independent of perception and intuition. The notion of proper justification provides a second sense in which foundational propositions might be said to be epistemologically privileged; they may be privileged, not in the sense that they are especially certain, but in the sense that they provide the real basis in terms of which other truths are justified. Hence, even if logic is not a certain foundation for arithmetic, it may nonetheless provide the proper justification for arithmetical truths. In this section, I want to examine this further presupposition of Frege's epistemology; the idea that there is such a thing as the proper justification of any given proposition.

In the *Grundlagen*, Frege is quite clear that he is not interested in how we actually come to know arithmetical truths, but in exhibiting the real justification for accepting them. In §17, he quotes Leibniz, "...it is here a matter, not of the history of discoveries, which is different in different people, but of the connection and natural order of truths, which is always the same." Here we come across a rationalist strand in Frege's thought; the idea that there is a 'natural order of truths', where this implies that there must be such a thing as *the* real or true ground for a proposition. This assumption is clearly implicit in Frege's classification of truths as either *a posteriori* or *a priori*, analytic or synthetic; which of these terms applies to a judgement depends on the nature of the first principles appealed to in *the* proof of the proposition.[40] In introducing these distinctions, Frege writes:

> It not uncommonly happens that we first discover the content of a proposition, and only later give the rigorous proof of it, on other and more difficult lines...In general, therefore, the question of how we arrive at the content of a judgement should be kept distinct from the other question, Whence do we derive the justification for its assertion?....When a proposition is called a posteriori or analytic in my

[40] Michael Dummett makes just this point when he remarks that "[t]he basis of [Frege's] classification is the justification for the judgement: not how we in fact know the proposition to be true, but the best justification of it that *could* be given." [Dummett 1991, p.23].

> sense, this is not a judgement about the conditions, psychological, physiological and physical, which have made it possible to form the content of the proposition in our consciousness; nor is it a judgement about the way in which some other man has come, perhaps erroneously, to believe it to true, it is a judgement about *the ultimate ground upon which rests the justification for holding it to be true.*
>
> [Frege 1884, §3 my emphasis]

There are two distinctions in the background here which need to be separated. The first is the distinction between *discovering* a truth and having some *evidence* for it. Carl Hempel gives a nice example of this distinction:

> The chemist Kekulé, for example, tells us that he had long been trying unsuccessfully to devise a structural formula for the benzene molecule when, one evening in 1865, he found a solution to his problem while he was dozing in front of his fireplace. Gazing into the flames, he seemed to see atoms dancing in snakelike arrays. Suddenly, one of the snakes formed a ring by seizing hold of its own tail and then whirled mockingly before him. Kekulé awoke in a flash; he had hit upon the now famous and familiar idea of representing the molecular structure of benzene by a hexagonal ring. He spent the rest of the night working out the consequences of this hypothesis.
>
> [Hempel 1966, pp. 15-16]

A famous anecdote of Poincaré's provides a mathematical example:

> Just at this time, I left Caen, where I was living, to go on a geologic excursion under the auspices of the School of Mines. The incidents of the travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go to some place or other. At the moment when I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time, as, upon taking my seat in the omnibus, I went on with a conversation already commenced, but I felt a perfect certainty. On my return to Caen, for conscience's sake, I verified the results at my leisure.
>
> [Poincaré 1907]

In both cases, someone made a discovery, or at least first came to believe a proposition, without having any real evidence for it. The evidence was found later; by

experiment in the case of benzene, proof in the case of the Fuchsian functions; only then was the discovery confirmed. There is therefore a more or less clear distinction between discovery and evidence. But we can make this distinction easily enough without assuming there is such a thing as the *proper* justification for a proposition. All we need is the idea that some processes of belief formation involve citing evidence for a proposition and others do not. It is not that in such cases we come to accept certain truths on the basis of inadequate evidence; evidence which does not provide the proper justification of the truths in question; Kekulé's dream and Poincaré's sudden insight do not provide any *evidence* for their hypotheses at all.

The second distinction however, brings us closer to the idea of proper justification. This is a distinction between different kinds of evidence we can have for a proposition. As Frege says, we can come to accept a proposition initially on grounds which we later come to realise are not the best grounds that could be given. Einstein, for example, made use of arguments based on verificationism (the view that the meaning of an empirical statement consists in the method by means of which we would verify it) in some expositions of the special theory of relativity.[41] The opinion of most scientists these days would be, I hope, that such arguments are irrelevant to the justification of the theory; the true test of the theory is its success in making empirical predictions. One might be tempted to say; it is the power of the theory to explain certain phenomena which provide the *real evidence* for it, not the philosophical arguments. In the same way, Frege argues that although we may initially come to accept such arithmetical truths as $7 + 5 = 12$ or the associative law of addition on empirical or inductive grounds, this cannot be the real evidence for those truths; the proper justification of the truths of arithmetic must consist in proofs of them from first

[41] See for example [Einstein 1917, p. 22].

principles. Likewise, Bolzano would have argued that the evidence for the intermediate value theorem provided by the intuitive geometric argument was specious; the real justification for that theorem was given by his analytic proof.

More generally, we may come to know a truth by hearing it from an authority. Most of our layman's knowledge of science and mathematics has this source; we come to know that the speed of light is constant for all observers by reading Einstein or other commentators on the theory of relativity; we come to know that Fermat's Last Theorem is true by reading in books or articles that Andrew Wiles proved it. We might say that in such cases, although we may have *some* evidence that a certain propositions is true, we may not know the *real* justification for it; namely the proof in the case of Fermat's Last Theorem or the experimental evidence in the case of the theory of relativity.

However, all we need in order to account for such examples is the idea that there can be better and worse kinds of evidence for a proposition. We do not need the idea that every proposition has a proper justification – the best justification of it that *could* be given. The fact that there are better and worse kinds of evidence for a hypothesis is well known in the natural sciences. A statistical correlation between A and B provides some evidence that there is a causal connection between A and B; if we can explain the causal mechanism involved by means of a well established theory, we have a better kind of evidence. That an established theory predicts the existence of some new elementary particle provides some evidence that the particle exists; an experimental observation of the particle is better. More controversially, some say that if a theory predicts something already known, then this provides some evidence for the theory, but a prediction of some new and unexpected phenomenon is a stronger kind of evidence for the theory.

In mathematics too, there can better and worse kinds of evidence. We can have inductive evidence for a mathematical conjecture for example, by confirmation of a large number of its instances; a deductive proof is a better kind of evidence. On the other hand, some proofs are said to be more explanatory and hence provide better evidence for a theorem than others.[42]

We can say then, that the evidence for the special theory of relativity which comes from the power of that theory to explain certain phenomena is stronger than the evidence (if such it can be called) that comes from verificationism. In the same way, although some evidence for numerical equations and arithmetical laws can be obtained from induction or from successful applications, Frege's proofs may provide better evidence for them, since they go some way to explaining *why* they are true, rather than merely providing some evidence *that* they are true. Likewise, the intuitive argument for the intermediate value theorem provides some evidence for it, but Bolzano's analytic proof provides a better kind of evidence, since it gives a better explanation of the theorem. We can say this without committing ourselves to the claim that there could not be any better explanation of it, nor do we need to suppose that there must be such a thing as the *best* explanation of it.

It is a commonplace that evidence comes in degrees; some evidence for a proposition can give us stronger grounds for accepting it than other evidence. It does not follow from this that there is such a thing as the *best*, ultimate or fundamental grounds for accepting it.[43] That claim provides one way of giving sense to the doctrine of foundationalism; the foundations of a theory just are those propositions which state the true grounds for accepting that theory. If there is such a thing as *the* ground or justification

---

[42] This idea is discussed in more detail in chapter six.
[43] Of course, there may be such a thing as the best justification for a statement that we can give at any particular time, but this obviously does not mean that we might not find a better justification in the future.

for mathematical statements, then mathematics has foundations and it may be the task of the philosopher to uncover them. But if there is no reason to believe there is such a thing as *the* grounds for accepting a proposition, then mathematics has no such foundation.

6. Mathematics Without Foundations

The same point was made, in a different way, by Hilary Putnam, in 'Mathematics without Foundations' [Putnam 1975d]. In that paper Putnam argued against the view that there is such a thing as the foundation for any mathematical theory. He argued that any mathematical theory may be given many different, but equally explanatory, equivalent formulations:

> In my view, the chief characteristic of mathematical propositions is the very wide variety of equivalent formulations that they possess. I don't mean this in the trivial sense of cardinality: of course, every proposition possesses infinitely many equivalent formulations; what I mean is rather that in mathematics the number of ways of expressing what is in some sense the same fact (if the proposition is true) while apparently nor talking about the same objects is especially striking.
>
> [Putnam 1975d, p. 45]

Putnam gives an example from physics. We can describe a quantum system either as a system of particles or as a system of waves (in the rather strange quantum mechanical sense of 'wave' and 'particle'). The two descriptions are equivalent; each can be translated into the other and they have exactly the same empirical content. But the metaphysical pictures underlying the two descriptions are very different.

Putnam then describes two equivalent, but metaphysically very different formulations of arithmetic. The first formulation is the standard Dedekind-Peano axiomatization. Here we have quantification over abstract particulars; the natural numbers.

Putnam refers to this kind of description of mathematical reality as *Mathematics as Set Theory* – "the conception of mathematics as the description of a universe of mathematical objects" - platonism in other words. The alternative formulation he refers to as *Mathematics as Modal Logic*. In the case of arithmetic, the idea here is to consider every arithmetical proposition as making a claim about what must hold good in any ω-sequence, that is, in any structure isomorphic to that of the natural numbers.

Let AX be the conjunction of the second-order Dedekind-Peano axioms for arithmetic. Let A be any statement of arithmetic expressed in the same language. If A is true, then it is a second-order logical consequence of the Dedekind-Peano axioms. That is A is true just in case the implication: AX → A is valid, that is iff:

(1) $\Box(\mathrm{AX} \rightarrow \mathrm{A})$

The formulas AX and A will still contain the undefined terms, *zero*, *successor* and *natural number*. Replace these terms wherever they occur in AX and A by the second-order variables S, o and N. We write the result as AX(S,o,N) and A(S,o,N) respectively. The formula AX(S,o,N) says that the relation S, the object o and the property N have a certain mathematical structure – that of any ω-sequence. The formula A(S,o,N). says that A is true in the structure formed by S, o and N. The proposed translation of A is then:

(2) $\Box\forall S\forall o\forall N[\,\mathrm{AX}(S,o,N) \rightarrow \mathrm{A}(S,o,N)\,]$

This says that necessarily, if S, o and N constitute an ω-sequence, then A is true in that ω-sequence. If A is true, then (2) will be a valid formula of second-order modal logic.[44]

The effect of such a translation scheme, Putnam argues, is to show that every statement of arithmetic, platonistically interpreted, has an equivalent formulation as a statement which tells us what must of necessity hold good in any ω-sequence. The metaphysical picture of this latter formulation is quite different to the platonistic one. According to the modal-logic picture, arithmetic is not a theory about particular abstract objects; the natural numbers. Rather it is the theory of what must be true in any structure of a certain kind. We have in effect, quantification over *structures*; more precisely, over properties and relations; but no quantification over individual numbers. Putnam writes:

> ...each of these two ways of looking at mathematics can be used to clarify the other. If one is puzzled by the modalities .... then one can be helped by the set-theoretic notion of a *model* (necessity = truth in all models; possibility = truth in some model). On the other hand, if one is puzzled by the question recently raised by Benacerraf: how numbers can be 'objects' if they have *no* properties except order in a particular ω-sequence, then, I believe, one can be helped by the answer: call them 'objects' if you like (they *are* objects in the sense of being things one can quantify over); but remember that these objects have the special property that each fact about them is, in an equivalent formulation, simply a fact about *any* ω-sequence. 'Numbers exist'; but all this comes to, for mathematics anyway, is that (1) ω-sequences are *possible* (mathematically speaking); and (2) there are *necessary* truths of the form 'if α is an ω-sequence, then....' (whether any concrete example of an ω-sequence exists or not).
>
> [Putnam 1975d, p. 49]

[44] In Putnam's paper, the example he gives of this translation scheme is first-order, rather than second-order. I have used a second-version of the translation scheme here for ease of exposition. The problem with the first-order version is that we cannot take the conjunction of the first-order Dedekind-Peano axioms as the antecedent of our conditional. This is because there are infinitely many such axioms, in particular, infinitely many instances of the induction schema. For more details on this and other structuralist formulations of arithmetic see Parsons [1990].

Geoffery Hellman, in his book *Mathematics Without Numbers* [Hellman 1989] developed this idea of Putnam's as the basis for his *modal-structuralist* account of mathematics. His hope was to avoid the problems associated with platonism by giving an account of mathematics in which quantification over abstract objects was eliminated. But this was not quite Putnam's aim. For Putnam, the two formulations, platonist and modal-structuralist stand on an equal footing; neither is a more or less correct description of mathematical reality than the other, in just the same way as both the wave and particle descriptions of the quantum system are equally correct. Putnam's point was that we have here two equally explanatory 'foundations' for arithmetic. We cannot say of either that it provides *the* ground or foundation for arithmetic. *The* implies uniqueness and in the case of mathematics (as in the case of the empirical sciences) uniqueness is just what we do not have.

I think that Putnam's metaphysical point is correct and that we can draw an epistemological conclusion. We have two equally correct, but metaphysically quite different descriptions of mathematical reality. Since there are equally correct, they must be on equal footing epistemologically, as well as metaphysically. It cannot be that our evidence for one of them is better or worse than the evidence for the other, just as it cannot be that we have better or worse evidence for one rather than another of the equivalent formulations of quantum mechanics. Since mathematics has no unique metaphysical foundation, it cannot have a unique epistemological foundation either. Both of the formulations Putnam describes provide an equally good explanation of the truths of arithmetic, although the explanations provided are quite different. Hence, each provides an equally good *justification* for arithmetic.

We might want to say; the real justification for a proposition *p* must appeal to whatever it is that makes *p* true – a *truth-maker* for *p* if you like. Perhaps this is why we want to say that when I know Fermat's Last Theorem by hearing it has been proved from an authority, I do not know the real reason why it is true; for my evidence is based on facts about the beliefs of various mathematicians and not on whatever it is that makes Fermat's Last Theorem true. But if there is no such thing as *the* correct metaphysical description of mathematical reality, but many equally correct such descriptions, then there is no unique answer to the question of what makes a particular mathematical proposition true. Hence, even if to give a real justification for a proposition, one must say something about the facts that make it true, there will be still be no unique real justification for that proposition. I could justify a mathematical proposition *p*, by citing something that makes it true, but some other justification, in terms of some other, equally correct description of what makes *p* true would do just as well.

We have to give up the idea that there is such a thing as *the* justification for a mathematical proposition. Mathematical propositions can be justified in various ways; some as good as each other, some better or worse than others. In mathematics, there are no epistemological foundations, only better and worse kinds of evidence.

CHAPTER TWO

POST-FOUNDATIONAL EPISTEMOLOGY

Since the collapse of the great foundational programmes, attention has shifted away from epistemology and towards metaphysical issues; over the last few decades there has been a proliferation of proposed ontologies for mathematics. Nonetheless, epistemological issues have not been completely ignored. As we shall see, a great deal of this metaphysical work can be seen as motivated by epistemological concerns. The problem is no longer seen to be one of providing an *a priori* guarantee that the mathematics we accept is true, but rather one of reconciling the ontology of mathematics with an empirical or naturalistic account of human knowledge. Although foundationalism has been largely abandoned, many philosophers continue to see the role of epistemology in this area as normative; the problem is to provide an account of the subject matter of mathematics in terms of which it is possible to give a broadly empirical justification for our mathematical beliefs. In this chapter, I want to examine and criticise certain aspects of these 'post-foundational' accounts of the epistemology of mathematics. I shall begin by discussing an argument which has provided one of the main sources of motivation for this work.

1. BENACERRAF'S DILEMMA

By the mid-twentieth century, views on what was still often called 'the foundations of mathematics' could be divided into three main kinds, each offering a distinct analysis of the concept of mathematical truth. A central debate was between *platonists* and *constructivists*. According to the platonist position, mathematical truth depends on the existence of abstract objects; a typical variety of this account elaborated it by arguing that all of classical mathematics can be developed from within the theory of sets.[1] In opposition to this was the intuitionist or constructivist account of mathematical truth, according to which a mathematical statement can be said to be true (or false) only when it is possible to construct a proof (or disproof) of it. Opposed to both of these were various kinds of *conventionalist* accounts of mathematical truth; typical was the view according to which mathematical statements are true or false in virtue of the meanings of their component terms – that is, *analytic*, in the new sense given to that term by the logical positivists.[2]

In a highly influential paper 'Mathematical Truth', Paul Benacerraf argued that *none* of these accounts of mathematical truth can be correct. [Benacerraf 1973]. According to Benacerraf, any adequate analysis of the concept of mathematical truth must satisfy two conditions. The first condition is that the analysis must be in conformity with a general account of the semantics of our language. Mathematical statements should not be treated differently to statements of other kinds having the same logical form; mathematical and non-mathematical statements alike should be treated in a uniform manner. Since to provide a semantics for a set of statements is to state the conditions under which they are true, this means that any adequate account of mathematical truth must conform to a general theory of truth for the sentences of our language. It is not enough for the analysis to simply label some statements with the tag 'true' and others with the tag 'false'. The analysis must also

[1] This view can be seen as one way of salvaging something from the logicist programmes of Frege, Russell and Whitehead; many took the view that what the work of these philosophers really showed was that mathematics was reducible not to logic, but to set-theory. See for example [Quine 1937, 1940].
[2] For the constructivist or intuitionist account see [Heyting 1931]. For very clear expositions of the conventionalist account see [Ayer 1936, Hempel 1945b].

explain why it is that those sentences labelled as 'true' can properly be said to be *true* and this can only be done in terms of a general account of the concept of truth.

The second condition is that any adequate analysis must be in conformity with a general theory of knowledge. Since some mathematical sentences are not only true, but known to be true, the account must explain how this is possible. That is, it must explain how we can know the mathematics that we do, in terms of a general account of the conditions required for knowledge.

Benacerraf then argues that no account of mathematical truth satisfies both of these conditions; all those accounts which satisfy the first condition fail to satisfy the second and conversely, all accounts which satisfy the second condition fail to satisfy the first:

> ...accounts of truth that treat mathematical and nonmathematical discourse in relevantly similar ways do so at the cost of leaving it unintelligible how we can have any mathematical knowledge whatsoever; whereas those which attribute to mathematical propositions the kinds of truth conditions we can clearly know to obtain, do so at the expense of failing to connect these conditions with analysis of the sentences which shows how the assigned conditions are conditions of their *truth*.
>
> [Benacerraf 1973, p. 662]

In order to establish this claim, Benacerraf divides accounts of mathematical truth into two broad categories. He asks us to consider the following two sentences:

(1) There are at least three large cities older than New York.

(2) There are at least three perfect numbers greater than 17.

These appear to have exactly the same logical form, namely:

(3) There are at least three *F*s that bear *R* to *a*.

If we analyse (1) as having the form (3), then according to Benacerraf, (1) will be true just in case the thing named by *a* (New York in this case) bears the relation *R* (*x* is older than *y*) to at least three things which satisfy the predicate F (x is a large city). In the same way, if we analyse (2) as having the form (3), then (2) will be true just in case the object referred to by a (the number 17) bears the relation R (x is greater than y) to at least three objects which satisfy the predicate F (x is a perfect number).[3]

Let us call any account which analyses both (1) and (2) as having the form given by (3), and which, in general, gives the same standard semantic analysis to mathematical and non-mathematical statements of the same form, a *standard* account. An example of a standard account would of course be the platonist analysis of mathematical truth. Many philosophers of mathematics of course, have denied that we should analyse mathematical statements in this way. An account which does not treat (1) and (2) as having the form given by (3), but which provides some alternative semantic analysis of mathematical statements we can call a *non-standard* account.[4] An example of a non-standard account would be the intuitionist analysis of mathematical truth, for on this account although the 'surface form' of (1) and (2) may be the same, the quantifiers and connectives in (2) are treated differently to those in (1); they are interpreted intuitionistically rather than classically. On this view (2) is not made true by abstract objects but our possession of a proof.[5] A more extreme non-standard view, which Benacerraf also mentions, would be the formalist account of Hilbert. As we saw in the previous chapter, on that account the

[3] A *perfect* number is one which is equal to the sum of its proper divisors. The smallest perfect number is 6; its proper divisors are 1, 2 and 3 and 1+2+3=6. The next three perfect numbers, all greater than 17, are 28, 496 and 8128.

[4] Benacerraf, somewhat misleadingly, calls them *combinatorial* accounts, since they often explain the meaning of mathematical statements in terms of purely syntactical features, such as provability.

[5] Of course, since 'x is a perfect number' is a decidable predicate, the intuitionistic and classical analyses of (2) would agree on the truth value of this statement, although not on its meaning.

semantic analysis of (2) would be radically different to that of (1), since (2) is a non-finitary or ideal statement and hence neither true nor false.[6]

Benacerraf now argues that his first condition on the adequacy of an analysis of mathematical truth rules out any non-standard account and that the second condition rules out any standard account. Hence no account is adequate. His argument for this startling conclusion depends on his elaboration of those two conditions of adequacy.

The first condition is that any adequate analysis of mathematical truth must conform to a general account of the semantics of our language; in particular it must conform to a general account of what it is for a sentence of our language to be true; a general theory of truth in other words. Benacerraf argues that our best general account of the concept of truth is that provided by Tarskian semantics:

> I take it that we have only one such account: Tarski's, and that its essential feature is to define truth in terms of reference (or satisfaction) on the basis of a particular kind of syntactico-semantical analysis of the language, and thus that any putative analysis of mathematical truth must be an analysis of a concept which is a truth concept at least in Tarski's sense.
>
> [Benacerraf 1973, p. 19]

Obviously, any standard account will satisfy the first condition, elaborated in this way. However, the condition rules out non-standard accounts, which give a different semantic analysis to mathematical statements. One problem with such accounts is that they prevent us from giving a uniform semantic analysis for our language taken as a whole. We have to treat mathematical and non-mathematical discourse as involving essentially different languages, with a corresponding distinction between different kinds of truth. For a

[6] The reason (2) is non-finitary is that its truth-value is not decidable in a finite number of steps; an algorithm which stepped through each number greater than 17, checked to see if they were perfect and halted when it had found three such numbers would fail to halt if the statement in question is false.

constructivist for example (at least one who restricts their constructivism to mathematics) the concept of truth as applied to mathematical statements will be quite distinct from that concept as applied to non-mathematical statements. There is a well known problem with the idea that there are different kinds of truth, which stems from the fact that we ought to be able to combine truths acquired from different fields of inquiry into a unified picture of the world. If we have a unified account of truth, independent of subject matter, then we can say for example that the conjunction of two statements 'A and B' is true just in case A is true and B is true. But if we have an account which analyses mathematical truth in a different way to truth as applied to, say, statements of physics, then we encounter a problem; if A is mathematically true and B is a truth of physics, then in what sense is 'A and B' true, if any? It is not a purely mathematical truth, nor a purely physical truth. If it is true in some more general sense, then we are back to the idea that there is such a general, topic neutral concept of truth. Perhaps it is not true at all then. But then we will be unable to explain how we can build up a coherent, unified account of the world, by combing knowledge obtained from different fields.

There is a further problem however. As already mentioned, it is not enough for an analysis of mathematical truth to simply show how we can label some sentences as 'true' and others as 'false'. There must be a reason to think that the concept characterised by the account is indeed a concept of *truth*. Benacerraf suggests that this can only be done by showing how the concept characterised as 'mathematical truth' is the same as the concept defined by our best general theory of truth. Since our best general theory of truth is provide by Tarskian semantics, any account which explicates mathematical truth along non-standard lines must fail to have explicated a concept of *truth*. Hence, the second condition,

elaborated in this way, rules out all non-standard accounts of mathematical truth. For example, the first condition requires that any account which equates truth in mathematics with *provability*, must explain the connection between provability and truth as defined by our best general theory of truth. But since our best definition of the concept of truth is Tarski's, it appears that any account of this kind will fail to satisfy the first condition, since provability is quite independent of truth so defined.[7]

The second condition on the adequacy of an analysis of the concept of mathematical truth is that it must explain how mathematical knowledge is possible. Analogously, Benacerraf argues that this can only be done by appealing to a general account of what it is to know something. In particular, we need to show how the conditions for knowledge required by such a general account can be satisfied in the case of mathematical knowledge.

Benacerraf elaborates this condition by arguing that out best general account of knowledge is some version of the causal theory of knowledge:

> I favour a causal account of knowledge on which for *X* to know that *S* is true requires some causal relation to obtain between *X* and the referents of the names, predicates and quantifiers of *S*... It must be possible to establish an appropriate sort of connection between the truth conditions of *p* (as given by an adequate truth definition for the language in which *p* is expressed) and the grounds on which *p* is said to be known.....In the absence of this, no connection has been established between *having those grounds* and *believing a proposition which is true*....The link between *p* and justifying a belief in *p on those grounds* cannot be made. But for that knowledge which is properly regarded as some form of justified true belief, then the link *must* be made.
>
> [Benacerraf 1973, pp. 23-24]

[7] In particular, the existence of undecidable statements of arithmetic (statements which are true, yet neither provable nor disprovable) appears to show that an account of this kind will violate Tarski's T-schema; for where A is such a statement A will be true, although 'A is true' will be false, since 'A is true' on such an account means 'A is provable'. As we shall see in section two however, this appearance is somewhat deceptive; it is open to the intuitionist to interpret the left hand side of the T-schema intuitionistically and if this is done the T-schema will remain valid.

But the second condition now rules out any *standard* account of mathematical truth. For on such an account, the truth of a mathematical statement requires reference to certain objects; numbers for example. Given that some mathematical statements are indeed true, this commits us to the existence of mathematical objects. But mathematical objects, if they exist, must be *abstract* objects. Although the distinction between abstract and concrete objects is difficult to make precise[8] it is easy to say what abstract objects are not. They are not physical objects, they are not located anywhere in space or time. They are not causal agents of any kind; they are not effected by, nor do they themselves effect, anything at all. But how is it possible to acquire knowledge of objects which cannot effect human beings or anything else in the world in any way? This is of course is the problem mentioned in chapter one, of explaining our knowledge of abstract objects; a problem which Frege attempted, but ultimately failed to solve.

Benacerraf conclusion is that the second condition rules out any standard account of the concept of mathematical truth: 'For a typical 'standard' account (at least in the case of number theory or set theory) will depict truth conditions in terms of objects whose nature, as normally conceived, places them beyond the reach of the better understood means of human cognition (e.g., sense perception and the like)' [Benacerraf 1973, p. 20].

In particular, since our best account of knowledge requires some causal connection between the knower and the objects which make the known statements true and since a standard account explains the truth-conditions of mathematical statements in terms of causally inert abstract objects, it follows that mathematical knowledge is in principle impossible according to such an account:

[8] See for example [Hale 1983].

> ....combining *this* view of knowledge with the 'standard' view of mathematical truth makes it difficult to see how mathematical knowledge is possible. If, for example, numbers are the kinds of entities they are normally taken to be, then the connection between the truth conditions for the statements of number theory and any relevant events connected with the people who are supposed to have mathematical knowledge cannot be made out. It will be impossible to account for how anyone knows any properly number-theoretic propositions. The second condition on an account of mathematical truth will not be satisfied, because we have no account of how we know that the truth conditions for mathematical propositions obtain.
>
> [Benacerraf 1973, pp. 24-5]

We are therefore faced with a dilemma. Either we give a uniform semantic analysis of mathematical statements along the lines of Tarski's analysis of truth, or we provide an alternative semantic analysis. If we take the latter course, we fail to satisfy the first condition – we will be unable to explain why some mathematical statements can be properly said to be *true*. If we take the former course, however, then although the first condition will be satisfied, the second condition will be violated – we will be unable to explain how some mathematical statements can be *known* to be true.

Of course there are many ways in which one might seek to avoid this dilemma.; much contemporary work in the philosophy of mathematics is usefully classified according to the way in which it attempts to solve it. One could, for example, accept a standard semantic analysis of mathematical statements and agree that some mathematical statements are true, but deny that this commits us to the existence of abstract objects, by providing an alternative account of the nature of mathematical objects, one which allows for a direct perceptual or causal connection between those objects and human beings. This is perhaps the most common type of solution to Benacerraf's dilemma. I examine some examples in section three, below.

On the other hand, one could accept the standard analysis and the truth claim, and also accept that this commits us to abstract objects, but deny that this makes mathematical knowledge impossible, by providing an account of how we can come to know truths about abstract objects without there being any direct causal connection between us and them. This is the route taken by Quine and Putnam; I will examine their arguments in chapter four. On the other hand, more radically, one could accept the standard analysis, and the claim that *if* some mathematical statements are true, this commits us to abstract objects, but deny the truth claim, holding that mathematical statements are all literally false. This is Hartry Field's solution to the problem, which is also discussed in chapter four.

All of these solutions to Benacerraf's dilemma retain something like the standard semantic analysis of mathematical statements, but seek in different ways to avoid the conclusion that this makes mathematical knowledge impossible. With one or two notable exceptions, the non-standard approach has been pretty much abandoned. One of the exceptions is of course, Michael Dummett who has argued for an intuitionistic interpretation of mathematics. Dummett rejects the assumption implicit in Benacerraf's argument, that the meaning of a sentence is to be analysed in terms of truth conditions along Tarskian lines. Instead he proposes that the meaning of a statement is to be analysed in terms of its assertability conditions; to know the meaning of a statement is to know what would count as verifying it. In the case of mathematical statements, this entails that to know the meaning of such a statement is to know what would count as a proof, or disproof of it. Dummett's main argument for this conclusion however, depends on very general considerations about the communicability of meaning; if valid his argument would apply to statements of any kind whatsoever and not just to mathematical ones. Hence Dummett can

avoid the problems associated with failing to give uniform semantic analysis of our language. Dummett can be seen as avoiding Benacerraf's dilemma by replacing the classical Tarskian conception of truth with an alternative general account of that concept. The problem of course is to say in more detail what this concept comes to and how it can be justified, a task which Dummett undertakes with great vigour in his writings on the subject. [see for example, Dummett 1975]. This solution however, comes at a price – many theorems of classical mathematics are disprovable when interpreted intuitionistically. As we shall see, this price is considered too high to pay by many contemporary philosophers of mathematics.

Benacerraf's dilemma can be stated in the following way. Our best general theory of truth, namely Tarskian semantics, combined with the thesis that some mathematical statements are true commits us to the existence of abstract mathematical objects. But our best general theory of knowledge, namely the causal theory, entails that we cannot have any knowledge of such objects and that mathematical knowledge is therefore impossible. The second part of Benacerraf's argument concludes that knowledge of abstract objects is impossible and many philosophers have taken something like this argument as a standard formulation of the epistemological problem for platonism; the problem of showing how we can acquire knowledge of causally inert, abstract objects. Many argue that this problem is indeed insoluble and that platonism should therefore be abandoned and replaced with an alternative ontology for mathematics.

There are however, reasons for dissatisfaction with Benacerraf's way of stating his dilemma; in particular with the way he elaborates the two conditions on the adequacy of accounts of mathematical truth. I shall argue that properly formulated, Benacerraf's dilemma is independent of both Tarskian semantics and the causal theory of knowledge. One upshot of this analysis will be that in fact, there is a more general epistemological problem in the philosophy of mathematics; a problem which is independent of *any* general account of knowledge and also independent of *any* theory of the subject matter of mathematics. I shall begin however, by examining the horn of Benacerraf's dilemma which proceeds from Tarski's theory of truth to the existence of abstract objects.

2. Tarskian Semantics

The first reason for dissatisfaction with Benacerraf's formulation of the dilemma is its dependence on what he calls 'Tarskian semantics' or 'Tarski's theory of truth'. The problem here is that Tarskian semantics is itself a *mathematical* theory. Let us start by reviewing the details of that theory.

In 'The Concept of Truth in Formalized Languages' [Tarski 1956] Tarski inaugurated what has become known as *model-theoretic semantics* by showing how to provide a semantics for the formal language of first-order predicate logic.[9] The key notion in Tarskian semantics is that of *truth in a model.* A *model* for a formal language L is a set-theoretic structure M consisting of several parts:

(1) A non-empty set D; the *domain* of M.

(2) A function that assigns to each singular term, or name, c of the language L an element c^M of D.

[9] See also [Tarski 1944].

(3) A function that assigns to each n-ary predicate symbol P of L a relation P^M on D. (That is, a set of ordered n-tuples of elements of D). If L has the identity symbol =, then $=^M$ is the identity relation on D.

(4) The set of truth values: $\{T, F\}$.

An *L-valuation based on the model M*, is then recursively defined to be a function v which maps elements of the language L onto elements of the model M in the following way:

(i) If x is any variable: $v(x) \varepsilon D$

(ii) If c is any singular term, or name: $v(c) = c^M$

(iii) If P is an n-ary predicate symbol and $t_1, \ldots t_n$ are terms (either variables or singular terms): $v(Pt_1 \ldots t_n) = T$ if $\langle v(t_1), \ldots, v(t_n)\rangle \varepsilon P^M$; $v(Pt_1 \ldots t_n) = F$ otherwise.

In addition, assuming our language contains only the two connectives $\rightarrow$ (if...then) and ~ (not...) and a single quantifier $\forall$ (for all...) then, if α and β are any well formed formulas:

(iv) $v(\alpha \rightarrow \beta) = T$ if $v(\alpha) = F$ or $v(\beta) = T$; F otherwise.

(v) $v(\sim \alpha) = T$ if $v(\alpha) = F$; F otherwise.

(vi) $v(\forall x\beta) = T$ if $v_{x/u}(\beta) = T$ for all $u \varepsilon D$; F otherwise.

where $v_{x/u}$ is the valuation based on M which differs from v only in that it assigns the element u of D to the variable x.

A valuation v is said to *satisfy* a formula α if and only if $v(\alpha) = T$. Tarski's formal definition of truth in a model can then be stated as follows:

(TT) A sentence of a formal language L (a formula of L containing no free variables) is *true in a model M* if and only if it is satisfied by every L-valuation based on M.

Suppose we are given a formal language powerful enough to express the Dedekind-Peano axioms for arithmetic Then we can define a model N for that language. The language will need two predicate symbols, = and S for the identity and the successor relations and just one name, 0 for the number zero. The domain of N is just the set of all natural numbers. S^N is the set of all ordered pairs $\langle a, b\rangle$ such that a and b are natural numbers and a is the successor of b. $=^N$ is just the identity relation on the natural numbers and 0^N is the number zero. By (TT) a sentence of this language is true in a model if and only if it is satisfied by every valuation based on that model. We might call a sentence of the language of arithmetic true *simpliciter* if it is true in the 'standard model' N, that is:

(TTA) A sentence of the language of arithmetic is *true* if and only if it is satisfied by every valuation based on the model N.

Consider for example, the first Dedekind-Peano axiom, which states that zero is not the successor of any number. In our formal language this sentence would be expressed as $\forall x{\sim}S0x$. By (TTA) this sentence is true if and only if for every valuation v based on N, we have $v(\forall x{\sim}S0x) = T$. Now:

$v(\forall x{\sim}S0x) = T$

iff	$v_{x/u}({\sim}S0x) = T$ for all $u \in N$	By (vi)
iff	$v_{x/u}(S0x) = F$ for all $u \in N$	By (v)
iff	$\langle v_{x/u}(0), v_{x/u}(x)\rangle \notin S^N$ for all $u \in N$	By (iii)
iff	$\langle 0^N, v_{x/u}(x)\rangle \notin S^N$ for all $u \in N$	By definition of $v_{x/u}(0)$
iff	$\langle 0^N, u\rangle \notin S^N$ for all $u \in N$	By definition of $v_{x/u}(x)$
iff	$\langle 0, u\rangle \notin S^N$ for all $u \in N$	By definition of 0^N

So our sentence is true if and only if, for every natural number u, the ordered-pair $\langle 0, u\rangle$ is not an element of the successor relation; that is, if and only if the number zero is the not the successor of any natural number.[10]

The point I want to draw attention to is that the Tarskian definition of truth (TT) is a *mathematical* definition of a *mathematical* property. A formal language is itself a kind of mathematical object and a model for such a language is a certain kind of mathematical structure. Truth in a model is defined using the concept of a mathematical function – a valuation v. A sentence of a formal language is true if every such function meets some complex condition. Arithmetical truth, on such an account has become a mathematical property of certain mathematical objects (sentences of a certain formal language) just as being composite or prime is a mathematical property of numbers.

If platonism is true, then of course, the mathematical claim that a sentence α is true in the model N commits us to the existence of various abstract objects – in particular to the

[10] We have here proved an instance of one half of Tarski's T-schema: 'α' is true $\leftrightarrow \alpha$.

model N and hence to all the natural numbers. But in this regard, the statement 'α is true in N' is no different to any other mathematical statement. For the platonist, any true mathematical statement commits us to the existence of those abstract objects which make it true. But of course, platonism is not the only option. If one thinks that mathematical statements are in some sense true, but not in the platonist sense, then one ought to have the same attitude to Tarskian semantics itself. If for example, one adopted the intuitionistic analysis of mathematical statements, then one ought apply the analysis equally to Tarski's mathematical definition of the concept of 'truth in a model'. There is nothing to prevent the intuitionist from giving exactly the same recursive truth definition as Tarski, but interpreting the clauses intuitionistically rather than classically. Such an approach would yield all the instances of Tarski's T-schema; 'α' is true $\leftrightarrow \alpha$, where the right hand side of the equivalence is interpreted intuitionistically.[11]

Then again, suppose one had a *fictionalist* account of mathematics, according to which mathematical objects such as numbers do not exist at all, but nonetheless, mathematical statements can be true or false in the same sense (whatever that is) in which statements occurring in fiction can be true or false. Applying this to the definition of truth in model, the claim that every valuation based on N satisfies the sentence α (the claim that α is arithmetically true) makes a claim about non-existent, fictional objects. On this interpretation, one can say that 'α is true in N' without being committed to the existence of any mathematical objects.[12]

[11] See [Dummett 1975, p. 80].
12 One might of course be committed to fictions, texts, stories or whatever is involved in the analysis of 'fictional truth'.

Hence, one can accept the Tarskian definition of arithmetical truth (TTA) and the claim that some mathematical statements are true, without being thereby committed to platonism, for one need not interpret (TTA) platonistically.

Tarski's mathematical definition of truth in a model then, forces us to accept the existence of abstract mathematical objects only under a certain interpretation of mathematics. Thus Benacerraf makes a mistake when he argues that our 'best theory of truth', Tarskian semantics, commits us to platonism. If the first horn of Benacerraf's dilemma is to have any force, we need to find a way of stating the argument which is independent of Tarskian semantics, for so elaborated, the first condition will *not* rule out non-standard accounts of mathematical truth such as intuitionism or fictionalism.

What we need to do is distinguish the mathematical theory of truth in a model from what might be called *referential semantics*. The central idea of the latter is that the truth value of a sentence is a function of the references of its component parts. The notion of reference employed here is a philosophical one; it denotes some kind of intensional relationship between words and the world. Notice that no such relationship is required in the mathematical definition of truth in a model; all we have there are mathematical functions from one kind of mathematical object (items in a formal language) to others (items in a certain kind of mathematical structure). No substantial notion of reference is involved here.

It is referential semantics, rather than the account of truth in a model, which I suspect Benacerraf really has in mind when he speaks of our 'best theory of truth'. On such an account, the truth value of an arithmetical statement is a function of the referents of its component parts. The connection with Tarskian semantics is that one can use the recursive clauses (i)-(vi) as a mathematical model of the way in which the semantic values of the components of a sentence combine to yield the semantic value of the whole sentence. So '3 > 2' for example is true if the object referred to by '3' stands in the relation referred to by '>' to the object referred to by '2'. Benacerraf's argument is really that what this shows is that for '3 > 2' to be true, the symbols '3' and '2' must refer to objects. Such objects would of course be numbers. So referential semantics, given that some arithmetical statements are true, appears to commit us to the existence of numbers.

Contrast this referential account of mathematical truth to the logical positivist analysis of mathematical truths as analytic; where analytic statements are 'true in virtue of meaning'. On this account, the truth of '3 > 2' does not involve reference to any peculiar mathematical objects like numbers, rather it is made true by the stipulations we have laid down governing the use of the symbols '3', '2' and '>'.

A referential semantics for arithmetic then, commits us to the existence of *numbers*, but of course, it does not by itself imply that numbers are *abstract* objects. Again, further philosophical argument is needed to get from referential semantics to platonism. Benacerraf is well aware of this; his argument is just that the *usual* standard account of the mathematical objects required to make mathematical statements true is that such objects are abstract, rather than physical or mental entities. If so, we then have the problem of explaining how there can be knowledge of such objects.

3. Methodological Platonism

However, even this reformulation of the first horn of Benacerraf's dilemma will not quite do. To see this, consider that an obvious solution to the dilemma formulated in this way would be to retain a referential semantics for mathematical language, but to reinterpret mathematical sentences so as to avoid commitment to abstract objects, by providing an alternative, non-platonist account of the ontology for mathematics. The further hope is that we will then be able to explain our mathematical knowledge in terms some kind of normal epistemic contact (such as perception) with the modified ontology.

In her highly stimulating article 'Philosophy of Mathematics: Prospects for the 1990s', Penelope Maddy discusses several account of this kind.[13] For example, structuralists like Resnik and Shapiro give an interpretation of mathematics according to which it is concerned not with particular abstract objects, but with the properties of *patterns* or *structures*.[14] Hellman adopts a similar position, but develops the ideas of Putnam mentioned in chapter one, to give us a *modal-structuralist* interpretation of mathematics; on this account mathematics is about possible structures and what must of necessity hold good of them.[15] Then again, some writers have adopt a broadly *constructivist* interpretation of mathematics; Charles Chihara for example, provides a reformulation of mathematics in terms of the construction of open sentence tokens[16] while Philip Kitcher presents the theory of the *ideal collector*, according to which mathematics is concerned with the operations of collecting and correlating objects which could be carried out by an idealised

[13] See [Maddy 1991].
[14] See [Resnik 1981, 1982, Shapiro 1983].
[15] See [Hellman 1989].
[16] See [Chihara 1990]

agent.[17] Maddy herself takes a different approach, arguing that some sets are in fact not abstract, but *material* objects and hence that mathematical sentences about them are made true or false by physical objects we can directly perceive.[18]

I will discuss the question of whether these accounts succeed in their epistemological aims in section five below. Here, I want to draw attention to a feature shared by all the accounts of mathematics in Maddy's list. This is that they involve a refusal to take mathematical discourse at face value. Consider again the first Dedekind-Peano axiom:

(1) Zero is not the successor of any natural number.

On the face of it, (1) is of the form:

(2) $\sim\exists x(\mathrm{N}(x) \,\&\, \mathrm{S}0x)$

Following Benacerraf, let us call (2) the *standard* interpretation of (1). On most of the accounts mentioned above, the standard interpretation of (1) is rejected. On these accounts (1) does not have the form (2) but something quite different. For example, on the modal-structuralist account, as we saw in chapter one, the form of (1) would not be (2) but:

(3) $\Box\forall S\forall o\forall N\,[\,\mathrm{AX}(S, o, N) \rightarrow \sim\exists x(N(x) \,\&\, Sox)\,]$

Of course, it is possible to give a non-platonist interpretation of mathematics without claiming that the logical form of mathematical sentences is something other than the form they appear to have. One could claim for example that (1) does indeed have the form expressed by (2) but that the quantifiers of (2) range not over abstract objects, but over objects of some other kind; physical sets perhaps.

[17] See [Kitcher 1984]. Kitcher's theory is discussed in more detail in chapter three.
[18] See [Maddy 1990].

The problem with these accounts which I want to draw attention to here is that they are false to actual mathematical discourse. Contemporary mathematical discourse is at least methodologically platonist. It is the standard interpretation of the form of mathematical statements that mathematicians actually use. Mathematicians do talk about the subject matter of the fields in which they work in terms of certain objects – the natural numbers, sets and so on. Actual proofs proceed in terms of the standard interpretation of the form of mathematical statements and make use of principles of classical logic. Nor do mathematicians treat their objects as though they were physical. That mathematical objects are abstract is in a sense implicit in the practice of mathematics.

In actual practice then, mathematicians are platonists in their methodology, in the way they actually do mathematics. I take it that it would be preferable to have a philosophy of mathematics which explains this practice, rather than ignoring it or explaining it away. Otherwise, we have not given an account of mathematics as it is actually done. We will have, perhaps, a philosophy of a possible form of mathematics, not an accurate philosophy of actual mathematics.

If this is right then the first horn of Benacerraf's dilemma is quite independent of any semantic theory or general account of the concept of truth. We do not need referential semantics to yield the conclusion that mathematics is about abstract objects. All we need is the methodological principle that our account of mathematics should respect actual practice in mathematics. In particular, we should take mathematical discourse at face value. Consider an analogous case. It does not take any deep philosophical theory of semantics to tell us that nuclear physics is about sub-atomic particles. All we need to do is look at what physicists say and take what they say literally and at face-value. If we do this, we find that the claims they make concern the properties of various objects; protons, neutrons, electrons and so forth. Nor do we need any high powered philosophical argument in order to say something more about these objects; in particular that they have certain properties (mass, location in space and time) and lack certain others (colour, smell and so forth). Simply by paying attention to what physicists do and do not say about these objects, we can get to the conclusion that they are *physical* objects of a certain kind.

Exactly the same considerations apply to mathematics. We do not need referential semantics to tell us that mathematics is about certain objects. All we need to do is look at what mathematicians actually say. What they say are things like 'Every set has a power set' and 'Every number has a successor'. Taking this literally and at face-value, we conclude that mathematics is about certain objects; numbers, sets and so on. Again, it does not require much argument to conclude that these objects have certain properties and lack others. In particular, we need no deeply philosophical argument to the conclusion that these objects are not physical things, but *abstract*. All we need to do is look at what mathematicians do and do not say about their objects. Of course it is no *theorem* of mathematics that numbers do not exist in space or time or that they are not causal agents. But then neither is there any explicit law of nuclear physics to the effect that electrons occupy space and have mass but do not have colours or smells. These facts about electrons would be readily accepted, but are usually just taken for granted. They are implicit in the questions physicists do not ask and the things they do not say about their objects. The same goes for mathematics; that numbers are abstract is revealed by the questions mathematicians do not ask about them and the things they do not say about them.

If we take mathematical discourse seriously and at face value then, platonism of some kind or other seems forced on us. The reasons why we ought to take mathematical statements at face value are methodological, rather than dependent of any theory of the semantics of our language. The methodological principle involved is that we should take actual mathematical practice seriously. Platonism is to some extent implicit in this practice.

Now of course, this is not enough to rule out of court all of the non-platonist accounts mentioned above. There can be good reasons for denying that the face value or obvious interpretation of a kind of sentence is the correct interpretation. Attempts to solve philosophical problems have very often proceeded by arguing that certain kinds of statement ought not to be taken as having their face-value logical form. Russell's theory of descriptions is a famous example.[19] More mundane examples are provided by sentences like 'His whereabouts are unknown' or 'I did it for her sake'. The grammatical 'surface form' of such sentences is misleading as to their logical form; taking them literally and at face value would suggest that they are about strange objects such as *sakes* and *whereabouts* and of course they are not. In these and other cases, we get into some kind of philosophical trouble if we take certain sentences at face value; the trouble is resolved by showing that the statements involved have a quite different logical form to that which might be suspected by their surface form.

In assessing an argument of this kind we need to weigh up the cost of abandoning a literal, face value interpretation of some part of our language (the cost of claiming that we do not quite mean what we say) against the benefits, measured in terms of an ability to resolve philosophical problems, of the proposed reinterpretation of that part of the language. It may happen that the benefits do not in fact outweigh the costs. The new

[19] See [Russell 1905, Quine 1948].

interpretation may not in fact solve the problems it is intended to solve or it may raise new problems of its own. Then again, it may be that there was only an illusion of a real problem in the first place.

Applying this to the non-platonist accounts of mathematics listed above, we can say that it is a disadvantage of these sorts of accounts that they are false to actual mathematical discourse. I do not want to say that such accounts are therefore completely misguided; the point is just that the fact that they reinterpret mathematical discourse in a way that is at odds with actual mathematical practice is something that counts against them.

Of course, the claim would be that this disadvantage is outweighed by the metaphysical or epistemological advantages of such accounts. It can be argued however, that there is often very little such gain to be had; we get rid of numbers for example, but replace them with abstract structures or possible worlds, which raise new problems of a similar kind to those supposed to accrue to the platonist ontology.

The main advantage of these non-platonist accounts is supposed to be that they can avoid the epistemological puzzle about abstract objects; the problem how we can know anything about them. In order to properly assess these accounts of mathematics then, we need to get clearer about what exactly this problem amounts to. This brings us to the second horn of Benacerraf's dilemma and to the second reason for dissatisfaction with his formulation of it, namely its dependence on the causal theory of knowledge.

4. The Causal Theory of Knowledge

The causal theory of knowledge emerged from a recognition that justified true belief is insufficient for knowledge. The so called *Gettier cases* describe situations in which even

though someone has a true and justified belief that *p*, we do not want to say that they really *know* that *p*.[20] For example, suppose I read in the paper one morning that John Howard has resigned as Prime Minister of Australia. I read the same news in other papers, hear of it on the radio and see reports of the event on the television. I come to believe that John Howard has resigned and my belief is well justified. In fact, I have been made the victim of an enormous practical joke; all these news reports have been fabricated. Nonetheless, unbeknownst to the fabricators, John Howard has in fact resigned. So I have a justified true belief. But do I *know* that John Howard has resigned? Apparently not, for after all, I am only right by accident.

What seems to be missing in this case and others like it, is the lack of an appropriate causal connection between the formation of the belief and what makes that belief true. In the above example, my belief that John Howard has resigned does not count as knowledge because it was not caused by his resignation, but by the machinations of the fabricators. An obvious response to counter examples of this kind then, is to add a causal condition to the justified true belief account of knowledge; knowledge is justified true belief where there is an appropriate causal connection between the belief and what makes that belief true. This is still quite vague, since it has not yet been made clear what an 'appropriate' causal connection is. But there is no reason to suppose that the condition could not be clarified to some extent, at least in cases of perceptual knowledge.[21]

However, it is possible to describe Gettier cases in which it does not seem to be the lack of any appropriate causal connection which differentiates knowledge from justified true belief. Suppose I am sitting in a meadow, looking at some daffodils. I form the belief

[20] See [Gettier 1963].
[21] See for example [Gettier 1963, Goldman 1967 and Harman 1973].

that there are some daffodils before me. My belief is true and I have a perfect perceptual justification for it; it is a clear, sunny day and I am looking right at the daffodils. But I do not *know* that there are daffodils before me. What I am unaware of is that someone has filled this meadow with fake plastic daffodils, practically indistinguishable from the real thing. Although in this case my belief is correct and well justified, I am again only right by accident – if I had looked in a slightly different place, I would have been looking at a fake and still formed the belief that there were daffodils before me.

In cases like these, a person has a justified true belief and an ideal casual relation to the objects which make that belief true. But the knowledge claim can be defeated by facts of which the person is not currently aware and such that, if they were to become aware of them, they would be forced to retract their claim to knowledge. If I was informed that my meadow contained many fake as well as real daffodils, I would retract my claim to *know* that there were some daffodils before me – I would no longer be sure, even though in this case I happen to be right. Notice that the same kind of account can be given of why we do not have knowledge in the first kind of Gettier case. In the first example, if I became aware that I was the victim of fabricators, I would retract my initial claim to have *known* that John Howard had resigned, even if I learned that he had in fact resigned.

Hence there is another class of counter-examples to the justified true belief account of knowledge. In the first class, what seemed to be missing was an appropriate causal connection. But in the second class, something else seems to be missing; knowledge of some other facts which would defuse the knowledge claim. If we take all the counter examples now available to us into consideration, they do not point quite so obviously to the

causal theory. It is much less clear that an appropriate causal connection must be present for there to be knowledge. In this way, the motivation for the causal theory is undermined.[22]

The view of many philosophers these days seems to be that the causal theory cannot be defended as a thesis about necessary and sufficient conditions for knowledge. If this is so, then the causal theory poses no real threat to platonism.[23] Nevertheless, epistemological worries about platonism have persisted. The challenge still appears to have some force; if abstract mathematical objects are so severely independent of human beings, it does appear at least *obscure* how we can know anything about them. The problem has changed from a perception of an inconsistency between platonism and the causal theory of knowledge, to a challenge to provide a broadly empirical or naturalistic account of our knowledge of mathematical objects. W.D Hart put the point like this:

> It is a crime against the intellect to try to mask the problem of naturalizing the epistemology of mathematics with philosophical razzle-dazzle. Superficial worries about the intellectual hygiene of causal theories of knowledge are irrelevant to and misleading from this problem, for the problem is not so much about causality as about the very possibility of natural knowledge of abstract objects.
>
> [Hart 1977, pp. 125-6]

Several philosophers noting the contemporary ambivalence towards the causal theory of knowledge, but still suspecting that there is an epistemological problem for platonism have attempted to find a formulation of the problem which is independent of the causal theory of knowledge. For example, Hartry Field in the introduction to his book *Realism, Mathematics and Modality* argues that the way to understand Benacerraf's

---

[22] See also [Wright 1983, pp. 84-97].
[23] See for example [Steiner 1973,1975].

challenge to platonism is not as a challenge to justify our mathematical beliefs, but as a challenge to explain the *reliability* of those beliefs.[24]

According to Field, if we cannot produce an account of the mechanisms which explain how our mathematical beliefs so closely reflect the facts about mathematical objects, then this tends to undermine belief in those objects. Field attempts to express the idea that our mathematical beliefs reflect the mathematical facts without reference to 'facts' or to any notion of truth stronger than a disquotational one. The platonist must explain why it is that the conditional:

(CP) If mathematicians accept that '*p*', then *p*

holds for a wide range of mathematical statements *p*.[25] Penelope Maddy arrives at a similar conclusion:

> Even if reliabilism turns out not to be the correct analysis of knowledge and justification, indeed, even if knowledge and justification themselves turn out to be dispensable notions, there will remain the problem of explaining the undeniable fact our expert's reliability. In particular, even from a completely naturalized perspective, the Platonist still owes us an explanation of how and why Solovay's beliefs about sets are reliable indicators of the truth about sets.
>
> [Maddy 1990, p. 43]

Both Field and Maddy think that the platonist will be hard pressed to find such a solution. Field writes:

[24] See [Field 1989, pp. 25-30 and 230-9].
[25] [*ibid.* p. 26].

> ... to make it believable that the Benacerrafian challenge is insurmountable, one would have to argue that it is impossible to explain the reliability claim in question: one would have to argue that the various facts about how the platonist conceives of mathematical objects collectively rule out all possibility of finding such an explanation. (The relevant facts .. include their mind-independence and language-independence; the fact that they bear no spatio-temporal relation to us; the fact that they do not undergo any physical interactions ... with us or anything we can observe ...). Like Benacerraf, I refrain from making any sweeping assertion about the impossibility of the required explanation. However, I am not at all optimistic about the prospects of providing it.
>
> [Field 1989, p. 27]

In similar vein, we have Maddy:

> Obviously, what we are up against here is another, less specific, version of the same vague conviction that makes the causal theory of knowledge so persuasive: in order to be dependable, the process by which I come to believe claims about *x*s must ultimately be responsive in some appropriate way to actual *x*s. ...nothing can be responsive to non-spatio-temporal, unchanging, acausal, unobservable Platonic entities. How then can Solovay's reliability be anything more than a fluke? How can it possibly be explained?
>
> [Maddy 1990, p. 44]

Of course, the problem is not confined to platonism; Benacerraf's claim was that *any* philosophical account of mathematics has to explain how mathematical knowledge is possible. Instead of elaborating this as a requirement that we show how mathematical knowledge can be explained in terms of the causal theory of knowledge, Field and Maddy elaborate it by asking for an explanation of the reliability of our mathematical beliefs. Presumably any account of mathematics must provide such an explanation; the further suggestion is just that this problem will be especially acute for the platonist.

One thing to notice about this way of stating the epistemological horn of Benacerraf's dilemma is that it is independent of any general account of knowledge. We are not being asked to show how the conditions laid down in some analysis of '*X* knows that *p*' can be met in the case where p is a mathematical statement. Instead we are asked only to give an explanation of how we can have reliably true mathematical beliefs. However, there are some difficulties even with this reformulation of the problem. To make this clearer, we need to return to our examination of those accounts of mathematics which attempt to avoid Benacerraf's dilemma by providing an alternative, non-platonist interpretation of mathematical discourse.

5. THE NEW CONSENSUS

The accounts listed by Penelope Maddy (section three above) are all very different in their characterisations of mathematical reality. Nonetheless, Maddy argues that this work reveals a new consensus in the philosophy of mathematics:

> The new consensus as I see it, is this: some form of ontological tinkering can defuse Benacerraf's dilemma without sacrificing standard mathematics. That is to say, there are ways of understanding the subject matter of mathematics as something other than the cool and inaccessible inhabitants of Plato's heaven, as something accessible to human cognizers, without curtailing the practice of classical mathematics.
>
> [Maddy 1991, p. 156]

Although Resnik, Shapiro, Hellman, Chihara, Kitcher and Maddy disagree as to what kind of 'ontological tinkering' is required, they all more or less agree that a basic knowledge of their ontologies can be acquired empirically, by means of perception. For Resnik, simple facts about concrete patterns can be learnt perceptually.[26] Chihara's open sentence tokens, being physical inscriptions, are obviously perceivable. Kitcher explicitly appeals to perception as a means of giving us knowledge of some basic features of the

[26] [Resnik 1982, pp. 9-10].

activities of counting and correlating.[27] Hellman argues that we come to know that our mathematical structures represent a coherent possibility through the use of mathematics in empirically confirmed scientific theories.[28] Finally, Maddy gives a sophisticated psychological account of how our basic mathematical knowledge is acquired by perceptions of physical sets.[29]

Maddy's conclusion is that:

> .....the particular ontological details are less important than the general epistemological trend... there are two items of agreement: first that the traditionally platonist ontology of mathematics can be modified without sacrificing standard results and practices; and second that a very basic level of epistemic contact with the modified ontology is available through ordinary perception.
>
> [Maddy 1991, pp. 156-7]

In section three I argued that an adequate philosophy of mathematics should be consistent with actual mathematical practice. Part of that practice is obviously the language of mathematics and that language is at least implicitly platonist. So any account of mathematics which does not take mathematical language, as it actually is, at face value is though perhaps not inconsistent with mathematical practice, at least somewhat at odds with it.

The point I want to make here is that the same methodological argument also applies to the epistemology of mathematics. An adequate philosophy of mathematics should also be consistent with the ways in which mathematics is actually justified. The accounts in Maddy's list however, also fail to meet this condition. The problem is that perception can only give us knowledge of a restricted class of mathematical statements on

[27] [Kitcher 1983, pp 5, 117-118].
[28] [Hellman 1989, pp. 3, 96-7].
[29] [Maddy 1990].

such views as these – simple statements like '2 + 2 = 4' for example. Very little more advanced contemporary mathematics can be given a directly perceptual or empirical justification. How then, is the rest of mathematics known? How can our knowledge of the axiom of choice, or Fermat's Last Theorem, for example, be accounted for in terms of these reformulated ontologies? Most of these accounts fail to answer these questions, they provide no account of the ways in which the more theoretical portions of mathematics are actually justified.[30]

Again, I do not want to say that any of these accounts of the subject matter of mathematics are wrong. Rather, I want to take issue with Maddy's claim that 'some degree of ontological tinkering can defuse Benacerraf's dilemma'. In particular, these accounts do not completely avoid getting caught on the epistemological horn of that dilemma. It is not good enough to simply show that there is an interpretation of mathematical discourse under which we can have empirical or perceptual knowledge of mathematics For this tells us nothing about how mathematics actually is justified. Such an account need have nothing at all to do with the way mathematics is actually done.[31] We have to show not only that mathematical knowledge is *possible*, but how it is *in fact* obtained. Only in this way will we arrive at a solution to the epistemological problem which is properly sensitive to mathematical practice.

[30] Maddy and Kitcher are the two exceptions to this rule. On Maddy's account, for example, our most basic mathematical knowledge is grounded in our perceptions of sets of physical objects. Maddy then sets herself the task of explaining how the more advanced set-theoretic postulates (the axiom of choice, the continuum hypothesis and so forth) are justified. [Maddy 1988a, 1988b, 1990]. I will discuss some of the results of Maddy's investiagations into the justification of set-theoretic principles in chapter five. Kitcher's account of how mathematical knowledge grows out of rudimentary knowledge obtained via perception is examined in chapter three.

[31] As Philip Kitcher puts it: 'To show that it is in principle possible for us to have knowledge of mathematical reality... is not to explain how we do have such knowledge' [Kitcher 1984, p.103]. Here Kitcher is making the point with respect to platonism, but in fact it is generally applicable, to all accounts of mathematics.

We have seen then that the epistemological problem for platonism is not that knowledge of abstract objects is ruled out by the causal theory of knowledge. Field and Maddy argue that in fact the problem is independent of any theory of knowledge; it is to explain how we can have reliably true mathematical beliefs, in something like the sense of (CP) above. The problem with this is that an account can satisfy (CP) – explain why it is true – in a way that has little or nothing to do with the way in which mathematical beliefs are actually justified. Presumably perception is a reliable belief forming process and so if mathematical statements (suitably reinterpreted) can be known through perception, we will be able to explain the reliability of those beliefs. The foregoing remarks this is inadequate however, since the explanation offered has no obvious connection to the ways in which mathematical beliefs are *actually* acquired. If the problem is to explain the reliability of our mathematical beliefs, we would need to add a constraint on the possible explanations that we can accept. The explanation must make it clear how the actual practices of mathematicians can lead to reliable beliefs. We would need to explain why the conditional:

(CP_2) Actual mathematical practices are such that if a mathematician comes to accept '*p*' by means of them, then *p*

We can certainly say that explaining why (CP_2) holds for a large class of mathematical statements will be difficult for the platonist, but what is not clear is that the non-platonist accounts were are discussing do any better, to the extent to which they provide no account of actual mathematical practice.

Is the formulation of this problem given by (CP_2) the right way to make these worries clearer? One difficulty with it is that appears to ask us to explain the reliability of our justificatory practices in mathematics. This seems to be a request to justify our cannons of justification themselves. One might argue however, that this a request impossible to meet. By what standards of justification could we criticize or defend those standards themselves? Any such attempt would seem to face a problem of vicious circularity. This is of course, a version of Hume's problem of induction. Nelson Goodman has argued that it is indeed impossible to find a guarantee that our inductive practices are reliable. The same might be said for our justificatory practices in general, both inductive and deductive. For Goodman, the only real problem is that of simply describing, rather than attempting to justify our practice.[32]

I shall return to this issue briefly in the next chapter. Let us leave aside such worries for now. What I want to suggest is that there is a more fundamental epistemological problem for platonism. We should view the problem not as one of explaining how mathematical knowledge is possible in terms of a general theory of what it is to know something, nor should we see it as a problem of explaining the reliability of our mathematical beliefs. Rather, we should look at the problem in terms of a conflict between platonism and our normal everyday and scientific cannons of evidence and justification. Those cannons of justification are broadly empirical; we do not countenance as a source of evidence anything which cannot be somehow traced back to our normal perceptual contact with the world.

The real epistemological problem for platonism then, is to give an account of how there can be empirical evidence for the existence of such objects. But not just any account

[32] See [Goodman 1983].

will do; the explanation must also be consistent with the actual cannons of evidence in mathematics; it must show how evidential practice in mathematics can be seen to conform to the broadly empirical standards we make use of elsewhere.[33]

The difficulty seems to be that if mathematical objects are abstract, there can be no empirical evidence for their existence; but empiricism is the claim that all evidence is at root empirical.[34] Notice that this statement of the problem does not depend on any theory of knowledge, only the concept of evidence. The basic issue is independent of what (if anything) is required in addition to justified true belief for there to be knowledge. The problem is not really about knowledge, but about evidence; there can be no knowledge of abstract objects only if we can have no empirical evidence for them.

6. Restating the Problem

I have argued that Benacerraf's dilemma, properly interpreted is independent of any theory of truth or semantics and independent of any theory of knowledge. An overall condition on the adequacy of any philosophical account of mathematics is that it should be consistent with and hopefully explain mathematical practice. This condition has two aspects, corresponding to Benacerraf's two conditions on the adequacy of any account of mathematical truth. The first condition is that our account should accurately reflect the *subject matter* of mathematics, which means it must be consistent with the *language* of

[33] If we assume those general cannons are a reliable means of getting to the truth, then we will have gone about as far as we can go toward explaining the reliability of our mathematical beliefs.

[34] Quine's solution is to show that despite appearance, we can show how there can be empirical evidence for abstract objects. As we shall see in chapter four, however, Quine's account may also face the problem of failing to conform to mathematical practice; the kind of justification which Quine provides for mathematics, although it may be part of the story, appears to have little to do with the ways in which mathematics is actually justified.

mathematical practice. The second condition is that our account should accurately reflect the *epistemology* of mathematics; which means it must be consistent with the *evidential standards* of mathematical practice. The dilemma is then as follows. If our account of mathematics is to satisfy the first condition, we must take mathematical statements at face value. But on the face of it, mathematics is about abstract objects. Combined with the claim that some mathematical statements are true, the first condition tends to commit us to some form of platonism. But then we run the risk of failing to satisfy the second condition, since it seems as though it will be very difficult to give an account of how we can have *any* kind of empirical evidence for the existence of abstract objects, and *a fortiori* we will be unable to give any account of the evidence for them which is consistent with evidential practice in mathematics.

The first condition does not of course entail that no non-platonist account of the subject matter of mathematics can be correct. The condition gives us only a *prima facie* case for platonism, since the suggestion that we should take mathematical statements at face-value could be overruled by the disadvantages of doing so. However many of the accounts mentioned above also fail to satisfy the second condition. They show how there can be evidence, of a broadly empirical kind for mathematics, but they fail to do so in a way that is consistent with actual evidential practice in mathematics.

On many contemporary accounts of the epistemology of mathematics, the problem is still seen as one of *justification* – the problem is to show how there can be a perceptual or empirical justification for the mathematics we accept. That is to say, the approach to epistemology is still *normative*, rather than *descriptive*. Although we no longer expect to find an epistemologically secure and certain foundation for mathematics, we nonetheless

seek to find some sort of broadly empirical justification for our mathematical beliefs. I have argued that this will not quite do. We want our epistemology to be consistent with actual mathematical practice. This means that we should not be satisfied with an account which shows there is some way in which mathematics could be empirically justified – for such a justification might not have any connection at all with the ways in which mathematics is actually justified.

Many contemporary philosophers of mathematics have taken on board at least one aspect of my overall condition on the adequacy of an account of mathematics - that the account should be consistent with actual mathematical practice. Any account which abandons certain results of standard mathematics, or invalidates certain standard patterns of mathematical reasoning would obviously fail to meet this condition. One element of the new consensus identified by Maddy is indeed that any adequate philosophy of mathematics should not seek to curtail any of the results of classical mathematics. This has become a standard objection to intuitionism. Part of the reasoning behind it, stems I suspect, from a rejection of the ideal of 'First Philosophy'. The idea is that philosophy cannot provide us with an especially privileged standpoint from which the corpus of our ordinary scientific methods can be criticised. Scientific claims should be judged by ordinary scientific standards, philosophy can provide no higher standard than this. There can be no *better* reason for believing the truths of science than the reasons given by the usual scientific evidence for them.[35]

Applying this idea to the case of mathematics, it follows that mathematical claims ought to be judged by mathematical standards and that philosophy is not in a position to reject them on the basis of some supposedly higher standard. If this is right however, it

[35] See for example [Quine 1969a, 1975, Putnam 1971, pp. 71-4].

means that there can be no real problem of *justification* in the philosophy of mathematics. If philosophy cannot criticise, it cannot justify either. If mathematical claims are to be judged only by mathematical standards, then they cannot be justified in a way that goes beyond those standards. But to some extent, this is just what many contemporary accounts of the epistemology of mathematics attempt to do – they attempt to provide a justification for our mathematical beliefs which seems to have little connection with the standards of justification internal to mathematics itself.

Very few philosophers would think any more that theories in physics or chemistry stand in need of any kind of justification that cannot be provided by the methods of the physical sciences themselves. It is time to admit that the same is true of mathematics. Justifying mathematics is the task of the mathematicians; the task of the epistemologist is to provide a descriptive, systematic account of the ways in which this justification is carried out.

Even if there is a problem of justifying mathematics which properly falls to philosophy, the project of giving a descriptive account of mathematical practice will not go away. Clearly we will need to be able to give an accurate description of that practice, before we can even begin to attempt a justification of it. Furthermore, the problem of description is independent of any epistemological worries we might have about platonism. I argued above that the problem for the platonist is not just to show that there can be empirical evidence for abstract objects, but to show how the standards of evidence mathematicians actually employ can provide us with evidence for abstract objects in a way that is consistent with empiricism. Equally clearly, we need to be able to say what those standards of evidence are before we can assess the chances for a successful solution to this problem.

Indeed, the problems of descriptive epistemology are independent not only of platonism, but of any account of the subject matter of mathematics. I suggest that one way we can make some progress in the philosophy of mathematics is by carrying out an ontologically neutral study of the ways in which mathematics is justified. By concentrating on giving a purely descriptive account of the nature of mathematical evidence, we can properly ignore a great many questions about the subject matter of mathematics. We will find that many of the problems and questions which arise in the process of describing mathematical evidence are independent of any particular ontology for mathematics. Of course, ultimately we would like to be able to answer the ontological questions; we would like, if possible, a better understanding of what mathematics is about; the nature of the objects of mathematics. But a descriptive, ontologically neutral approach to epistemology holds out the hope that we can avoid the grip of Benacerraf's dilemma as I have stated it here, at least for a time. For on this approach, we can go a long way towards constructing an adequate epistemology for mathematics, one which accurately reflects mathematical practice, without first having to solve the ontological problem.[36]

The real problem in the epistemology of mathematics is not one of showing that the entire body of our mathematical knowledge can be built up on the basis of a sure and certain foundation. Nor, more generally, is it the problem of showing how our mathematical beliefs can be justified. Rather, the problem is to give a descriptive account

[36] Indeed, it may be that solving the epistemological problem as I have stated it here is a necessary step on the road towards solving the ontological problem. One might suppose that providing an accurate descriptive account of the nature of justification in mathematics would point the way to the correct view of the ontology of mathematics, since the way in which a subject is justified must have some connection to what it is about. However, the connections between the epistemology and ontology of mathematics are in fact quite loose; a descriptive epistemology for mathematics can be consistent with very many distinct ontologies for mathematics. Nonetheless, as we shall, an adequate account of the epistemology of mathematics can at least suggest that some accounts of the subject matter of mathematics are more explanatory of mathematical practice than others.

of the nature of evidence in mathematics. Our goal should be to describe mathematics in its own terms and seek to explicate the standards of evidence internal to mathematics itself. To do this, we need to pay much closer attention to the historical development of mathematics than is usual in philosophical accounts of the subject.

To make these ideas clearer, it will be helpful to look at some examples of the kind of approach to the epistemology of mathematics I have been arguing for. In the next chapter, I examine the work of two philosophers who have also argued that we cannot understand the growth of mathematical knowledge without understanding the real history of mathematics, and who have taken on the challenge of giving an accurate descriptive account of the epistemology of mathematics.

CHAPTER THREE

DESCRIPTIVE EPISTEMOLOGY

In the first two chapters I traced some of the history of the epistemology of mathematics. We have seen something of the evolution of a certain kind of normative approach to epistemological questions, beginning with the work of Frege, through the rise and fall of the three great foundational programmes to the modern 'new consensus'. I would like now to turn to a somewhat neglected area in the history of our subject, the evolution of an alternative descriptive approach to the epistemology of mathematics. I will discuss the work of two exponents of this alternative approach, Imré Lakatos and Philip Kitcher. I will be examining their work in some detail, for several reasons. Firstly, both writers attempt to develop the idea that mathematics is a science, albeit in different ways, so their work is directly relevant to my project here. Secondly, their work provides a paradigm example of the sort of descriptive, ontologically neutral approach to epistemology which I argued for at the end of the last chapter. Finally, and most importantly, the work of these writers contains many important insights concerning the nature of evidence in mathematics and the growth of mathematical knowledge.

1. LAKATOS'S PROGRAMME

After Frege's work, which attempted to unify mathematics and formal logic for the first time and the development of Frege's logicist programme along different lines by

Russell and Whitehead in *Principia Mathematica*[1], philosophical attention turned increasingly to the study of formal systems for mathematical theories. The development of intuitionism, Hilbert's programme and the philosophical impact of Gödel's incompleteness theorems gave added impetus to this research programme. By 1963, when Lakatos published *Proofs and Refutations* as a series of articles in the *British Journal for the Philosophy of Science*[2], the philosophy of mathematics was seen by many to be more or less identical with the study of mathematical logic. Lakatos thought that this development was harmful to the philosophy of mathematics. The introduction to *Proofs and Refutations* begins with the following words:

> It frequently happens in the history of thought that when a powerful new method emerges the study of those problems which can be dealt with by the new method advances rapidly and attracts the limelight, while the rest tends to be ignored or even forgotten, its study despised.
>
> This situation seems to have arisen in our century in the Philosophy of Mathematics as a result of the dynamic development of metamathematics. The subject matter of metamathematics is an abstraction of mathematics in which mathematical theories are replaced by formal systems, proofs by certain sequences of well-formed formulae, definitions by 'abbreviatory devices' which are 'theoretically dispensable' but 'typographically convenient'.[3]
>
> [Lakatos 1963, p. 2]

Lakatos thinks it is a mistake to identify mathematics with the model provided by formal axiomatic systems. Calling this identification 'formalism', he writes:

> ...there are problems which fall outside the range of metamathematical abstractions. Among these are all problems relating to informal ... mathematics and to its growth, and all problems relating to the situational logic of mathematical problem solving.
>
> [ibid. p. 2]

[1] [Whitehead and Russell 1910-13].
[2] [Lakatos 1963,1976].
[3] The quotations are from [Church 1956, pp. 76-77].

If we identify mathematics with formal systems, we will be unable to make sense of ordinary mathematical practice. For in practice, mathematicians do not set out their proofs as formal derivations from axioms, rather they present more or less rigorous arguments in a mixture of ordinary language and symbolism. The formal model has nothing to say about how such proofs are discovered - the 'logic of mathematical problem solving'. Nor does it tell us how the axiom systems were arrived at in the first place, how mathematical theories are created and justified, how they grow from previous mathematics. In general, the formal model provides no account of the process of mathematical discovery and hence it can make no sense of the real *history* of mathematics:

> Formalism disconnects the history of mathematics from the philosophy of mathematics...Formalism denies the status of mathematics to most of what has been commonly understood to be mathematics, and can say nothing about its growth. None of the 'creative' periods and hardly any of the 'critical' periods of mathematical theories would be admitted into the formalist heaven, where mathematical theories dwell like the seraphim, purged of all the impurities of earthly uncertainty.....Under the present dominance of formalism, one is tempted to paraphrase Kant: the history of mathematics, lacking the guidance of philosophy, has become *blind*, while the philosophy of mathematics, turning its back on the most intriguing phenomena in the history of mathematics, has become *empty*.
>
> [ibid. p. 2-3]

Lakatos then, is proposing a return to the study of real, informal mathematics - an examination of the methodology of mathematics, as revealed in its history. Lakatos had been deeply influenced by the work of Karl Popper in the philosophy of science.[4] Popper's target had been the inductivist account of the epistemology of science, according to which

[4] Indeed, the very title of Lakatos's work seems to be derived from Popper's *Conjectures and Refutations*, published the same year. [Popper 1963].

scientific laws are arrived at by a process of induction from observations. Popper rejected this view and proposed a different model. Scientific laws are proposed as tentative hypotheses. Scientific laws are not discovered by a process of induction from particular cases. They are arrived at in many and various ways, often just by guessing, but the way in which scientific laws are first discovered is really epistemologically irrelevant.

Once a law is proposed as a hypothesis, it must be tested against experience. This is achieved by deriving testable consequences from the hypothesis. The law L, implies some observation sentence O. We have $L \to O$. If O turns out to be false, we can validly infer that the law is false - if $L \to O$, then $\sim O \to \sim L$. The hypothesis has been tested and found wanting, it has been falsified. If the observation sentence O turns out to be true however, we cannot validly infer that the law L is true, this is just to commit the fallacy of affirming the consequent.

Nonetheless, if a law or theory entails many observation sentences, all of which turn out to be true and it entails no false observation sentences, then the law or theory has been 'corroborated'. It becomes accepted into the body of scientific knowledge. We have no guarantee, of course, that a well corroborated theory is true, only that it has not so far, been falsified. This leaves open the possibility that it will be falsified at some point in the future. If the theory is falsified, it is rejected and a new theory must be found to replace it. If the theory is corroborated, it is accepted, for the time being at least. The mark of a true science, according to Popper, is that its hypotheses are falsifiable in this way. He rejects as pseudo-scientific any body of propositions which cannot be falsified.[5]

[5] See opper 1959, 1963].

Lakatos wants to show that mathematics is a science in Popper's sense. In particular, he wants to show that mathematical claims are proposed as hypotheses which can be tested and hence falsified. Mathematics is not a body of sure and certain truths, it is fallible and revisable. This view, of course, runs against a tradition of long-standing in the history of the philosophy of mathematics, a view which Lakatos traces to what he calls 'dogmatism'. Lakatos aims to oppose the dogmatist view that mathematics is an infallible body of truths resting on sure and certain foundations with an alternative picture of mathematics - one more akin to Popper's picture of science - a picture in which mathematical claims are tentatively proposed as hypotheses to be tested, falsified and if necessary revised and improved. Lakatos does not directly argue for this position, instead he provides us with a case study - the history of a certain conjecture of Euler. In describing the history of this conjecture, Lakatos hopes to simply show, rather than directly argue for the superiority of his picture of mathematics over that of the dogmatists. He writes:

> The core of this case-study will challenge mathematical formalism, but will not challenge directly the ultimate positions of mathematical dogmatism. Its modest aim is to elaborate the point that informal, quasi-empirical mathematics does not grow through a monotonous increase of the number of indubitably established theorems but through the incessant improvement of guesses by speculation and criticism, by the logic of proofs and refutations.
>
> [ibid. p. 6]

The *method of proofs and refutations* is Lakatos's term for his model of the growth of mathematical knowledge. He presents his ideas in the form of a dialogue between a teacher and the students in a mathematics class. The discussion is concerned with Euler's conjecture that for all polyhedra:

$$V - E + F = 2$$

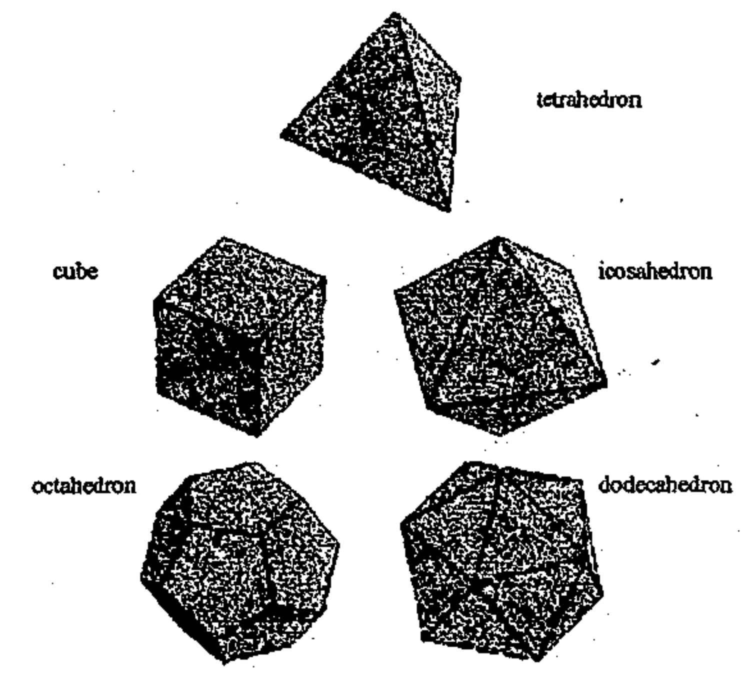
where V is the number of vertices, E the number of edges and F the number of faces of the polyhedron. As the discussion proceeds, the students recapitulate some of the history of the development of this conjecture – the real historical story is told in footnotes to the main text. By reflecting on the methods used to establish and improve the conjecture, the students state, reject and defend explicit methodological principles. The set of principles they finally arrive at is the method of proofs and refutations. This is meant not only as a description of how mathematics actually develops, but as a prescriptive heuristic for mathematical research.

2. The Method of Proofs and Refutations

2.1 A Conjecture and a Proof

The story begins then with Euler's claim concerning polyhedra. Euler had no proof of his conjecture, but he was nonetheless convinced that it was true. He pointed out that the conjecture is well corroborated by evidence which falls short of proof. It holds for all the regular polyhedra; the tetrahedron, octahedron, icosahedron, the cube and the dodecahedron (see figure 1). Euler also verified that the conjecture holds for a wide variety of other polyhedra; prisms, pyramids and so on. The conjecture also implies other known consequences, for example that there are only five regular polyhedra.[6]

[6] See [Euler 1750, Polya 1954a, pp. 35-43].



Polyhedron	*Vertices*	*Edges*	*Faces*
Tetrahedron	4	6	4
Cube	8	12	6
Octahedron	6	12	8
Dodecahedron	20	30	12
Icosahedron	12	30	20

Figure 1
The Regular Polyhedra

The teacher begins the discussion by offering the following proof of Euler's conjecture, originally due to Cauchy. Given any polyhedron, we remove a face and stretch the result so that it lies flat on the plane. Figure 2 shows the result of performing this operation on a cube. In doing this, the faces and edges may be deformed, but the number of edges and vertices will not alter. The number of faces however is decreased by one. Let Vp be the number of vertices, Ep the number of edges and Fp the number of faces of the resulting plane network. Then for the original polyhedron, we have:

(1) $V - E + F = Vp - Ep + (Fp + 1)$
$= (Vp - Ep + Fp) + 1$

The proof now proceeds by performing a sequence of operations on the plane network which do not alter the sum Vp – Ep + Fp. We start by triangulating the network. We draw diagonals on those polygons which are not already triangles. (see figure 3). In drawing a diagonal, we are adding an edge and a face. So Ep and Fp both increase by one. But if Vp – Ep + Fp = n for the original network, then

(2) $Vp - (Ep + 1) + (Fp + 1) = Vp - Ep - 1 + Fp + 1$
$= Vp - Ep + Fp$
$= n$

for the triangulated network. So triangulating the network does not alter the sum Vp – Ep + Fp. From the triangulated network, we remove triangles one by one. In removing a triangle, we either (a) remove one face and one edge (see figure 4a) or (b) we remove one face, two edges and a vertex (see figure 4b). In case (a) if Vp – Ep + Fp = n for the original network, then:

(3) $Vp - (Ep - 1) + (Fp - 1) = Vp - Ep + 1 + Fp - 1$
$= Vp - Ep + Fp$
$= n$

so the sum Vp – Ep + Fp remains the same. In case (b) if Vp – Ep + Fp = n for the original network, then:

(4) $(Vp - 1) - (Ep - 2) + (Fp - 1) = Vp - 1 - Ep + 2 + Fp - 1$
$= Vp - Ep + Fp$
$= n$

so again, the sum Vp – Ep + Fp remains the same.

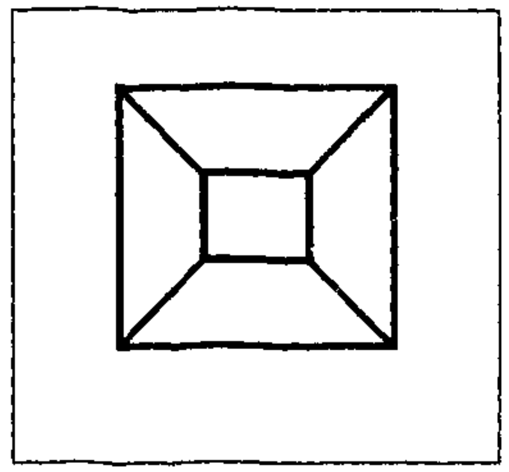


Figure 2

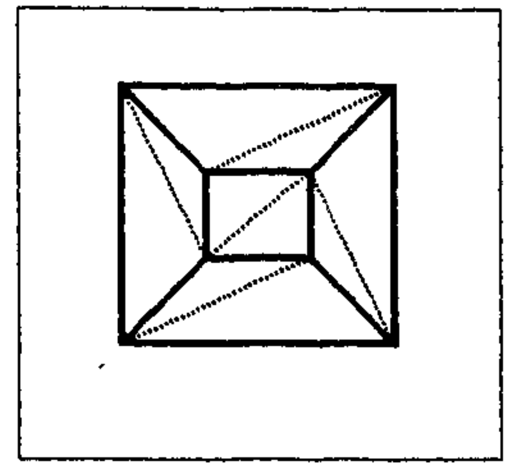


Figure 3

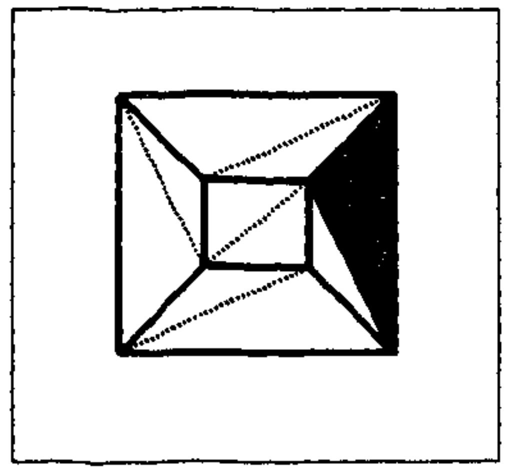


Figure 4a

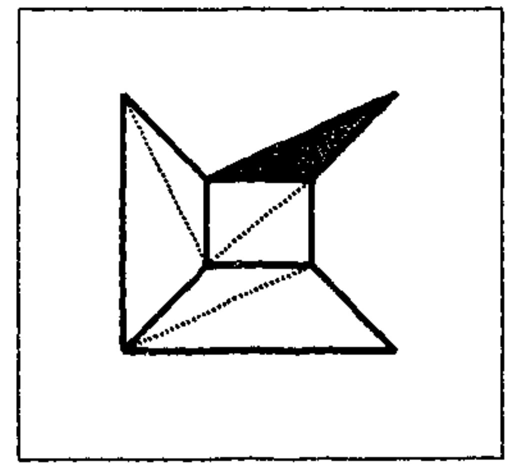


Figure 4b

At the end of this process, we get a single triangle. But for a single triangle, V – E + F = 1, since V = 3, E = 3 and F = 1. Since the sum Vp – Ep + Fp has not changed in the process of triangulating and then removing triangles, we must have:

(5) $Vp - Ep + Fp = 1$

for the original plane network as well as the final triangle. Substituting this value into equation (1), we obtain:

(6) $V - E + F = 2$

for the original polyhedron.[7]

The students immediately question this proof. They identify three main lemmas on which it depends. These are:

(1) Any polyhedra, after having had a face removed can be stretched flat on the plane.

(2) If we triangulate the result, we always get a single new face for every new edge we draw.

(3) In removing the triangles one by one, we either remove one face and one edge or we remove one face, two edges and a vertex.

The students express doubts about each of these lemmas. Notice what has happened here. Since Cauchy's proof is presented in a fairly informal way – the way in which most mathematical proofs are actually written – it is not immediately obvious what the logical structure of the proof is. The process of identifying the lemmas used in a proof and the way in which they entail the conclusion is called *proof-analysis* by Lakatos. If we were to attempt to formalise Cauchy's proof by translating it into the language of mathematical logic and devising a corresponding formal proof in an appropriate axiomatic system, proof-analysis would be our first step. Obviously this process is often neither automatic nor trivial. Furthermore, although a proof, once formalised, is either valid or not, there can be better and worse ways of analysing and so formalising the proof in the first place. As we shall see, the initial analysis of the proof that the students give is one that can be improved upon.

[7] See [Cauchy 1811, Lakatos 1963, pp. 8-9].

Our initially highly plausible conjecture is now seen to depend on at least three somewhat dubious lemmas. The proof then, seems to have actually *decreased* our level of confidence in the conjecture, rather than establishing it with certainty. The students naturally express some surprise and disappointment at this state of affairs. As the student Alpha remarks, 'But then we are worse off than before! Instead of one conjecture we now have at least three! And this you call a 'proof'!' [ibid. p. 10].

The teacher admits that what he calls Cauchy's 'thought-experiment' does not conclusively establish the truth of the conjecture. But if a mathematical proof does not show that its conclusion is true, what does it achieve? The teacher's response is as follows:

> ...I propose to retain the time-honoured technical term 'proof' for a thought-experiment – or 'quasi-experiment' – which suggests a decomposition of the original conjecture into subconjectures or lemmas, thus embedding it in a possibly quite distant body of knowledge.
>
> [ibid. p. 10]

Notice a consequence of this way of using the term 'proof'. A proof in this sense is not a guarantee of certain truth - in fact, it is not even a guarantee of truth at all. In this sense, it is possible to 'prove' a *false* conjecture:

> TEACHER: ...My interpretation of proof will allow for a false conjecture to be 'proved', i.e to be decomposed into subconjectures. If the conjecture is false, I certainly expect at least one of the subconjectures to be false. But the decomposition might still be interesting! I am not perturbed at finding a counterexample to a 'proved' conjecture; I am even willing to set out to 'prove' a false conjecture!
>
> [ibid. p. 25]

In fact, the third lemma is false as it stands. As one of the students points out, if we remove a triangle from *inside* the network shown in figure 3 instead of removing a boundary triangle, then we only remove a face - the edges and vertices remain unchanged. So if $Vp - Ep + Fp = n$ in the original network then:

$$\begin{aligned} Vp - Ep + (Fp - 1) &= (Vp - Ep + Fp) - 1 \\ &= n - 1 \end{aligned}$$

in the new network, so the sum $Vp - Ep + Fp$ has changed.

Notice that this is a counter-example only to the third lemma, it does not by itself refute the conjecture. Lakatos calls a counter-example to a lemma a *local counter-example*. A counter-example to the main conjecture, he calls a *global counter-example*. In this case, we have a local, but non-global counter-example. In response, the teacher suggests a modification of the lemma:

(3') In removing *boundary* triangles one, by one, we either remove one face and one edge or we remove one face, two edges and a vertex.

But there are counter-examples to this version of the lemma also. There is a way of removing boundary triangles one by one from the network, so that we end up with two disconnected triangles [ibid. p. 13]. The lemma is again reformulated:

(3'') There is always a sequence in which triangles can be removed one by one, so that at each stage the sum V-E+F does not alter.

Lemma 3 has now become quite a strong claim, lacking the initial plausibility of the original formulation.

We now come to the first *global* counter-example – the *hollow cube* (see figure 5). Think of this as a solid cube which has had a smaller cube hollowed out in its centre. This serves as a local counter-example to the first lemma, because it cannot be stretched flat on to the plane. But it is also a global counter-example, since for the hollow cube we have V = 16, E = 24 and F = 12 - double the number of vertices, edges and faces of a single cube. Hence, in this case, V - E + F = 4.

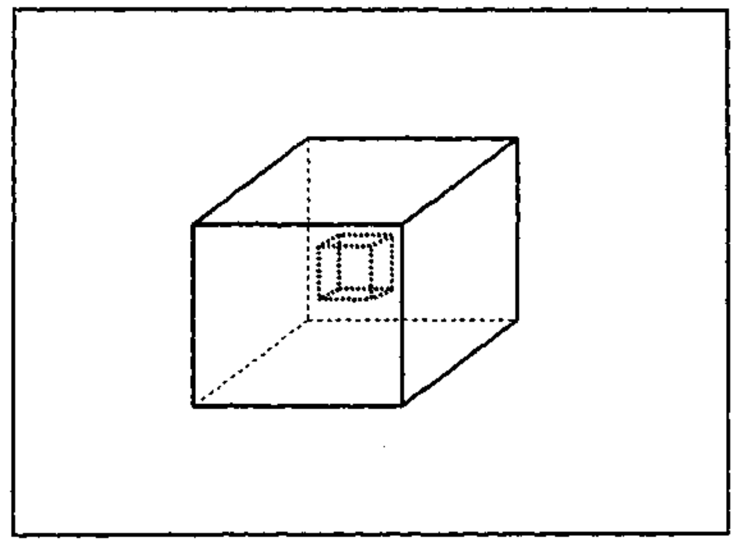


Figure 5
The Hollow Cube

2.2 THE METHOD OF MONSTER BARRING

We are now in a somewhat puzzling position. We appear to have both a proof of the conjecture *and* a counter-example to it. Given that Lakatos is using the term 'proof' to allow for even a false conjecture to be proved, this is not too surprising. But what should we do now? Should we simply reject the original conjecture? This is what Lakatos eventually suggests, but before doing so he canvasses the idea that we could instead *reject the counter-example*. The student Delta argues that the hollow cube is not a polyhedron, it is a *monster* and hence does not refute the conjecture that for all *polyhedra*, V – E + F = 2.

This leads to a definition of 'polyhedron' aimed at ruling out such monsters: *a polyhedron is a surface consisting of a system of polygons* [ibid. p. 13]. This definition rules out the hollow cube, since it consists of *two* surfaces, one inside the other.

There are counter-examples to the conjecture even under this interpretation of 'polyhedron' however. For example, there is the polyhedron, first discussed by Kepler, and named by him *the urchin* (see figure 6). This can be thought of as a polyhedron, in the sense of the definition given above, with twelve 'star-pentagon' (or pentagram) shaped faces (figure 7)[8]. This is a global counter-example to Euler's conjecture, since in this case, V=12, E = 30 and F = 12, so that V – E + F = –6.

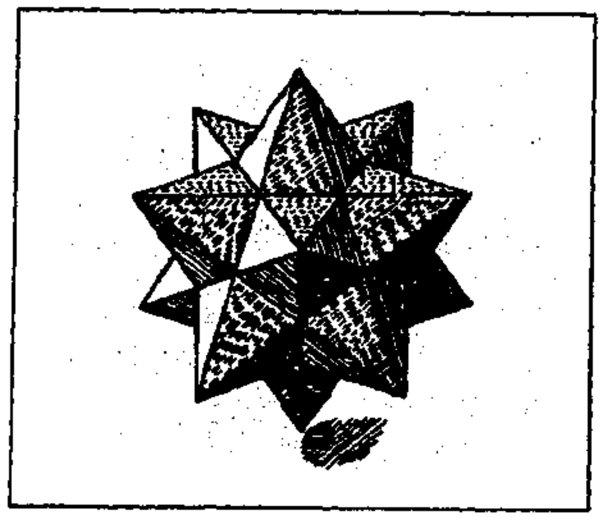


Figure 6
The Urchin

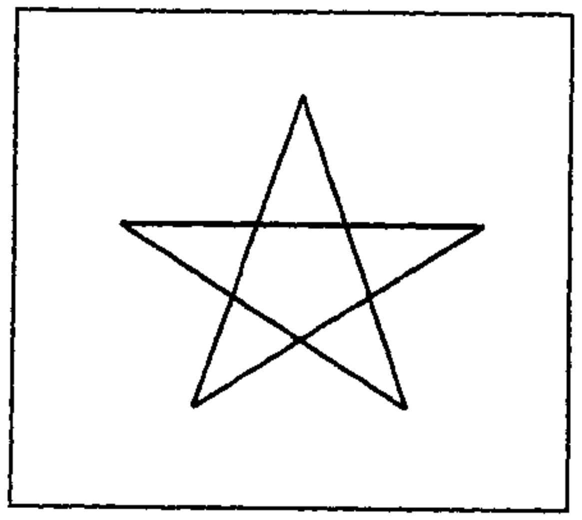


Figure 7
A star-pentagon face

But this counter-example only works if the pentagram counts as a polygon – for according to the current definition, a polyhedron is a surface consisting of *polygons*. Gamma suggests then, a definition of polygon which will rule this out: *a polygon is a system of edges arranged in such a way that exactly two edges meet at every vertex*. [ibid.

[8] Each of the twelve star-pentagonal faces are shaded differently in Kepler's drawing of the urchin shown in figure 6.

p. 19]. The pentagram is not a polygon in this sense, since it has *four* edges meeting at the vertices of the inner pentagon.

We are now introduced to yet another counter-example – the *picture-frame* (figure 8). This is a system of polygons in the sense of the new definition. However, for the picture frame, V = 16, E = 32 and F = 16, so V – E + F = 0 – another global counter-example to the conjecture.

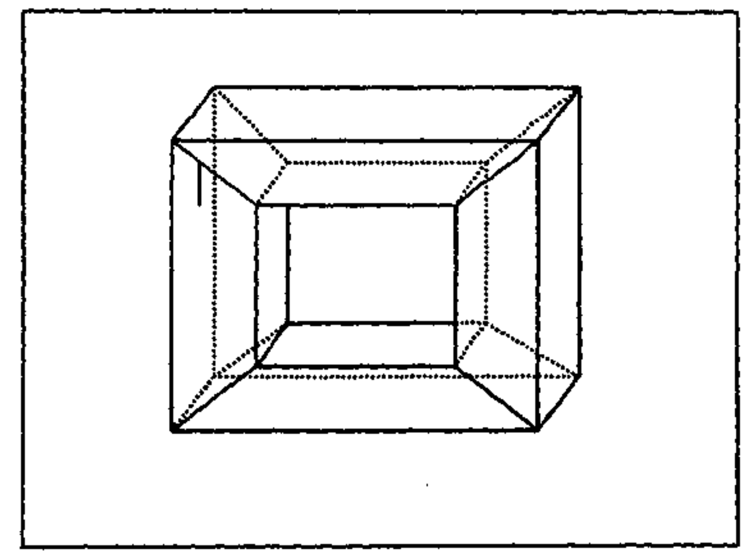


Figure 8
The Picture Frame

In response, an amendment is made to the definition of polyhedron: *for a genuine polyhedron through an arbitrary point in space, there will be at least one plane whose cross-section with the polyhedron will consist of a single polygon*. [ibid. p. 23]. In the case of the picture-frame, if we take a point inside the 'tunnel' and lay a plane through it, the intersection always consists of two distinct polygons. So this definition rules out the counter-example.

Lakatos refers to this method of saving the conjecture as *the method of monster-barring* and rejects it as *ad hoc*:

TEACHER: I think we should refuse to accept Delta's strategy for dealing with global counterexamples, although we should congratulate him on his skillful execution of it. We could aptly label his method *the method of monsterbarring*. Using this method one can eliminate any counterexample to the original conjecture by a sometimes deft but always *ad hoc* redefinition of polyhedron, of its defining terms, or of the defining terms of the defining terms.

[ibid. 25]

In what sense is this strategy *ad hoc*? Suppose we have a theory (or hypothesis) T and a piece of evidence E which is inconsistent with that theory (or hypothesis). That is, E is a counter-example to T. In response, we make a modification to the theory T, modifying it to T' say, so that T' is consistent with E. The modification to the theory is *ad hoc* just in case there is no *independent* ground for accepting T'. In other words, the modification is *ad hoc* if its *only* purpose is to save the theory from counter-example.

In this case we have the initial hypothesis that *for all polyhedra, $V-E+F=2$.* We discover a series of counter-examples to the hypothesis. The hypothesis is modified by redefining the terms polyhedron, polygon and so on, so as to exclude the counter-examples. But these modifications are *ad hoc*, because they often seem to have no independent justification - the only reason for accepting them is that they save the hypothesis from counter-example.[9]

[9] For example, in response to the picture-frame, the hypothesis is modified by proposing that for a genuine polyhedron, through an arbitrary point in space, there will be at least one plane whose cross-section with the polyhedron will consist of one single polygon. The picture-frame does not have this property, so the modified conjecture is immune to this counter-example. But what independent reason is there for accepting this amendment to the definition? If there is no independent ground for accepting it, other than that it saves the hypothesis from a counter-example, then we have a *ad hoc* modification of our hypothesis.

Other similar responses that reject the counter-examples are then suggested. For example, we could treat the counter-examples as *exceptions* to the theorem.[10] Alternatively, we can reinterpret the counter-examples so that they in fact support the conjecture rather than refute it.[11] The main problem with such methods, apart from their *ad hoc* character, is that they ignore the proof. They do not tell us what went wrong with the proof or how to improve it. The teacher proposes a new method. We should not reject the counter-examples as 'monsters' or 'exceptions'. Instead we have to accept that they refute the conjecture in its original form. But this is not the end of the story. By paying attention to the proof, we can use the counter-examples to *improve* our original conjecture.

2.3 The Method of Lemma Incorporation

The teacher begins by admitting that the picture-frame is a genuine counter-example to the conjecture. The picture-frame is a global and hence a local counter-example – it shows that one of the lemmas must be false. Which one? In this case, the guilty lemma is the first one – after removing a face, the picture frame cannot be stretched flat onto the plane. The teacher continues:

TEACHER: I accept the picture-frame as a criticism of the conjecture. I therefore discard the conjecture in its original form as false, but immediately put forward a modified, restricted version, namely this: the Descartes-Euler conjecture holds good for 'simple' polyhedra, ie for those which, after having had a face removed, can be stretched onto a plane. Thus we have rescued some of the original hypothesis. We have: *the Euler characteristic of a simple polyhedron is 2.* This thesis will not

[10] So for example, we think of the theorem as saying that 'for all polyhedra except those with tunnels, cavities, etc, $V-E+F=2$'.

[11] For example, instead of thinking of the urchin as having 12 star-pentagon faces, we can see it as having 60 triangular faces and hence 90 edges and 32 vertices. So now $V-E+F=32-90+60=2$ and the counter-example has become an example.

be falsified by the nested cube, by the twin-tetrahedra, or by the star-polyhedra - for none of these is 'simple'.

[ibid. 130-131]

The idea is that in the face of a global counter-example, we first identify the lemma or lemmas which are falsified by the example. We then restrict the domain of the main conjecture to that of the guilty lemma. In this case, we have a counter-example to the first lemma. So we restrict the conjecture to all those polyhedra for which the first lemma holds. Call any polyhedron which satisfies the first lemma *simple*. Then the revised conjecture is that for all *simple* polyhedra, $V - E + F = 2$. In this way, the basic structure of the proof is retained, but the conjecture is modified so as to incorporate the guilty lemma as a condition.[12] Lakatos calls this the *method of lemma-incorporation*.

A new global counter-example, the *crested cube* is now proposed. Here we have a small cube on top of a larger cube (figure 9). In this case, V=16, E=24 and F = 11, so $V - E + F = 3$. But the crested cube is a simple polyhedron – after having a face removed it can be stretched flat on to the plane.[13] So we have a global counter-example to our revised version of the conjecture.

[12] If we use Px for the predicate *x is a polyhedon*, Sx for *x is simple* and Ex for *x is Eulerian*, then we express the original conjecture as $\forall x(\mathrm{P}x \rightarrow \mathrm{E}x)$ (*All polyhedra are Eulerian*). We then discovered an object *a*, with the following properties:

Pa & ~Ea	*a* is a global counter-example to the conjecture
Pa & ~Sa	*a* is a local counter-example to lemma (1)

So we modify the conjecture to. $\forall x((\mathrm{P}x \,\&\, \mathrm{S}x) \rightarrow \mathrm{E}x)$(*All simple polyhedra are Eulerian*).Now *a* is no longer a counter-example to this revised conjecture, since we have (Pa & ~Sa) & ~Ea; for a counter-example to this improved conjecture, we would need (Pa & Sa) & ~Ea.

[13] If this is not clear, notice that the square which hides underneath the smaller cube is *not* a face of this polyhedron. So we can immediately stretch the top cube into a flat network that sits on top of the larger cube. After removing the bottom face from the larger cube, we can the stretch the entire network so it is flat on the plane.

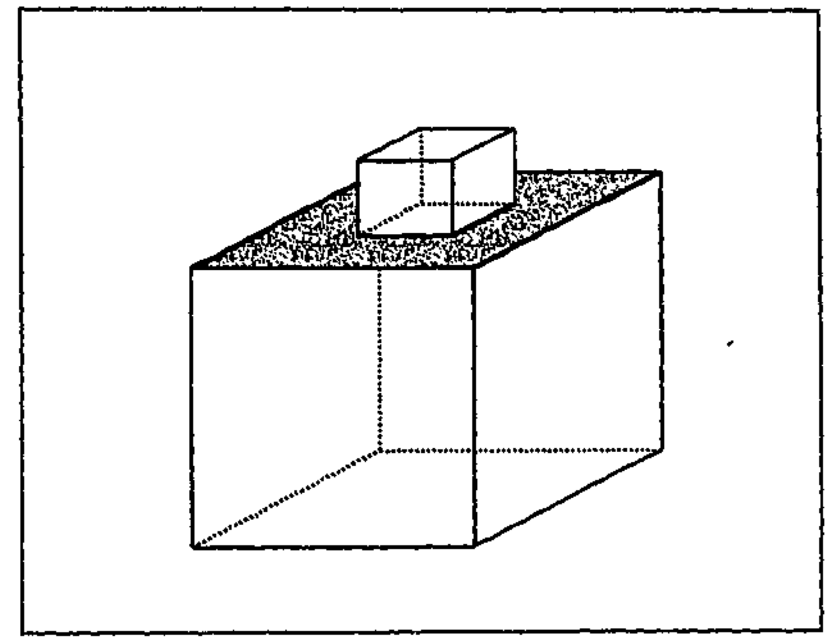


Figure 9
The Crested Cube

In response, the method of lemma incorporation is applied again. The crested cube is a counter-example to the second lemma, that in triangulating the faces, we always add an edge and a face. This is because the crested cube has a 'ring-shaped face' – the face of the large cube on which the smaller cube sits (shaded in figure 9). If we a draw a diagonal from a vertex of this face to a vertex of the smaller cube, we add an edge without adding a face [ibid. p. 132]. Hence the crested cube fails to satisfy a condition implied by the second lemma; that any face dissected by a diagonal edge falls into two pieces. Call a polyhedron which does satisfy this condition *simply-connected*. Then applying the method of lemma-incorporation, the new version of the conjecture is that *for all simple, simply-connected polyhedra*, $V - E + F = 2$.

The method of lemma-incorporation is sensitive to counter-examples which are both global (refuting the conjecture) and local (refuting one of the lemmas). What would happen if given a global counter-example, we cannot find a lemma which is refuted by it?

The method of lemma incorporation could not then be applied. Such a counter-example would be a *global*, but *non-local* counter-example. Are such counter-examples possible? How are we to deal with them?

Gamma points out that a *simple cylinder* is in fact a counter-example of this kind. The cylinder is not Eulerian – it has 0 vertices, 3 faces and 2 edges (the circles at the top and bottom), so in this case V – E + F = 1. Paradoxically, however, the cylinder is both simple and simply-connected, so it is not a local counter-example to any of the lemmas. This is odd because our proof seemed to show that any simple, simply-connected polyhedron *must* be Eulerian.

Why should we say that the cylinder is simple? If we remove the 'jacket' of the cylinder, it falls into two circular pieces. Nonetheless, the result can still be stretched flat onto the plane. We seem to have implicitly assumed in our proof that the resulting plane network should be connected – and this assumption fails for the cylinder.

Is the cylinder simply-connected? A polyhedra is simply connected if any face dissected by a diagonal falls into two pieces. The cylinder only satisfies this vacuously – the faces of the cylinder have no diagonals at all (since they have no vertices) and so there are no diagonals which do *not* cut a face in two. It seems that we implicitly assumed an existential reading of this condition; that any face when dissected by diagonal falls into two pieces *and there is at least one diagonal on each face*. Again, this condition fails for the cylinder.

An apparently global, but non-local counter-example really shows that there is something wrong with our proof-analysis. It shows that there may be hidden assumptions or lemmas in our proof, which we need to make explicit. Lakatos calls this the *principle of retransmission of falsity* – falsehood should be transmitted from the consequent of the theorem to its antecedent, the lemmas or conditions. Global, non-local counter-examples then, are a source of improved proof-analysis.[14]

On the other hand, improved proof-analysis can also lead to the discovery of new counter-examples. A proof may show how an initially plausible conjecture depends on lemmas which are questionable. Given a good proof-analysis, we may look for counter-examples to the lemmas (local counter-examples) and these may well also turn out to be counter-examples to the conjecture (global counter-examples). These are referred to as *proof generated refutations*. A proof brings to light lemmas which may lead us to counter-examples to the conjecture. Thus as Sigma remarks:

> SIGMA: Then not only do refutations act as fermenting agents for proof-analysis, but proof-analysis may act as a fermenting agent for refutations! ...
>
> LAMBDA: That is right. If a conjecture seems very plausible or even self-evident, one should prove it: one may find that it hinges on very sophisticated and dubious lemmas. Refuting the lemmas may lead to some unexpected refutation of the original conjecture....
>
> GAMMA: Then 'the virtue of logical proof is not that it compels belief, but that it suggests doubts'.[15]
>
> [ibid. 227]

Since proof and refutation are linked in this way, it is suggested that we should rename the method of lemma incorporation, the method of *proof and refutations*. The main aspects of this method are listed as three heuristic rules:

[14] Lakatos here adds an interesting historical footnote: 'Our class was rather an advanced one - Alpha, Beta and Gamma suspected three lemmas when no global counterexamples turned up. In actual history proof-analysis came many decades later: for a long period the counterexamples were either hushed up or exorcised as monsters, or listed as exceptions. The heuristic move from the global counterexample to proof-analysis - the application of the Principle of Retransmission of Falsity - was virtually unknown in the informal mathematics of the early nineteenth century.' [Lakatos 1963, p. 227]

[15] The quotation is from [Forder 1927, p. viii]. See also [Russell 1903, p. 360]: 'It is one of the chief merits of proofs that they instil a certain scepticism as to the result proved'.

Rule 1: Given a conjecture, set out to prove and refute it. Inspect the current proof to prepare a list of non-trivial lemmas (proof-analysis). Search for counterexamples to the conjecture (global counterexamples) and to the lemmas (local counterexamples).

Rule 2: If you find a global counterexample, discard the conjecture. Add to the proof-analysis a suitable lemma that will be refuted by the counterexample. Replace the discarded conjecture by an improved one that incorporates that lemma as a condition. Do not allow a counterexample to be dismissed as a monster. Make all 'hidden lemmas' explicit.

Rule 3: If you have a local counterexample, check to see if it is not also a global counterexample. If it is, apply rule (2).

[ibid. p. 229]

2.4 The Method of Proofs and Refutations

So far, we have considered both global and local counter-examples and global, non-local counter-examples. The former suggest improvements to the conjecture by lemma incorporation, the latter suggest improvements to the proof-analysis by making hidden lemmas explicit.

Omega points out that the method of proof and refutations, as described so far, has a certain flaw. By incorporating refuted lemmas into the proof as conditions we gradually restrict the domain of validity of the theorem. Our improved conjectures apply to a narrower and narrower domain of polyhedra. Hence, 'proof analysis, while increasing certainty, decreases content' [ibid. p. 236]. Lakatos refers to this as the *problem of content*.

Perhaps in our quest for certainty, we have 'withdrawn too far', leaving many Eulerian polyhedra outside the domain of our theorem. That is, we may have excluded many non-simple, non-simply connected polyhedra, which are nonetheless Eulerian. Such a polyhedron would represent a *local, non-global* counter-example to the theorem.

It is counter-examples of this kind which provide the impetus to increase content. Early on we found a counter-example to the third lemma. If we had invoked lemma incorporation at this point, our theorem would have applied only to the tetrahedron, for this is the only polyhedron for which one can remove triangles in any order at all without altering the sum $V - E + F$. Instead we modified the lemma so that it became immune to the counter-example. In this way, we avoided a drastic narrowing the domain of the theorem.

It is suggested that we add this as a heuristic rule to the method of proof and refutations:

Rule 4: Given a local counterexample which is not also a global counterexample, try to improve the proof-analysis by replacing the refuted lemma by an unfalsified one.

[ibid. p. 237]

It is pointed out that this rule has both a conservative and a radical interpretation. We can either replace the false lemma with a slightly modified one, keeping within the framework of the original proof. This is what was done in the case of the counter-example to the third lemma. Alternatively, we can *replace the false lemma, or even all the lemmas, by constructing a new, deeper proof.*

Some examples of different proofs are given. Omega gives a proof of Euler's conjecture due to Gergonne. The main lemma of this proof is that there is a face of the polyhedron which can be used as a camera lens to take a 'snapshot' of the interior of the polyhedron, from which all the edges and vertices are visible [ibid. p. 238]. Call any polyhedron which satisfies this condition a 'quasi-convex' or 'Gergonne' polyhedron. Then, after incorporating the lemma, we would have the theorem; *for all Gergonne* *polyhedra*, V – E + F = 2. Gamma points out that there are many simple polyhedra which although perfectly Eulerian are not Gergonne polyhedra. That is, there are many local, but non-global counter-examples to this theorem. Omega suggests we can imagine starting with Gergonne's proof, then finding that it is unsatisfactory by a local, but non-global counter-example - a polyhedron which is Eulerian but has no face from which every edge and vertex is visible. This is the sort of counter-example that Rule 4 looks out for. Such a counter-example would not of course, refute the theorem that 'all Gergonne polyhedra are Eulerian', but Rule 4 suggests that nonetheless some action is required – we should look for a deeper proof which uses new lemmas which are satisfied by the current counter-example. Instead of incorporating the false lemma (that all polyhedra are Gergonne) we replace this 'photographing' lemma by the wider topological or 'stretching' lemma. We thereby arrive at Cauchy's deeper proof and a wider theorem.

A proof due to Legendre is also mentioned [ibid. p. 239] but this proof applies to even fewer polyhedra than Gergonne's. Thus far Cauchy's proof is the deepest we have seen – it applies to the most polyhedra. But here too there are local, non-global counter-examples. One example is the *great stellated dodecahedron* (see Figure 10). This is a 'star-polyhedron' like the urchin (or *small stellated dodecahedron*). It too consists of star-pentagonal faces, but arranged in a different way. For this polyhedron, V=20, E=30 and F =12, so V – E + F = 2. This Eulerian polyhedron does not satisfy the first lemma however, since it cannot be stretched flat on to the plane. Another example is a slight modification of the picture-frame (see Figure 11). In this case, V=16, E=24 and F=10, so V – E + F = 2, but this polyhedron does not satisfy the first lemma either, since it is topologically equivalent to a torus, rather than a sphere.

These local, non-global counter-examples do not refute the theorem that 'all Cauchy polyhedra are Eulerian' (where a Cauchy polyhedron is one which is simple and simply-connected) but they suggest we should look for a deeper proof – one that will explain both Cauchy and Eulerian star-polyhedra for example.

One of the students claims to have found such a proof. This is in fact a proof due to Poincaré.[16] The teacher remarks that there is no time to discuss the proof in detail, he says only that it will not be the final word. If we state Poincaré's theorem as 'all Poincaré polyhedra are Eulerian' – after having incorporated all Poincaré's lemmas – then there will still be Eulerian, but non-Poincaré polyhedra – local, but non-global counter-examples.

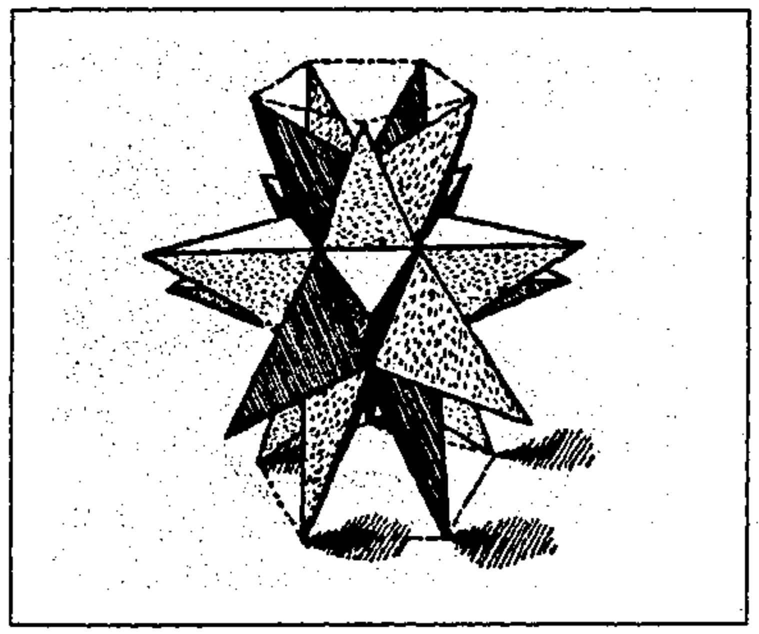


Figure 10
The Great Stellated Dodecahedron

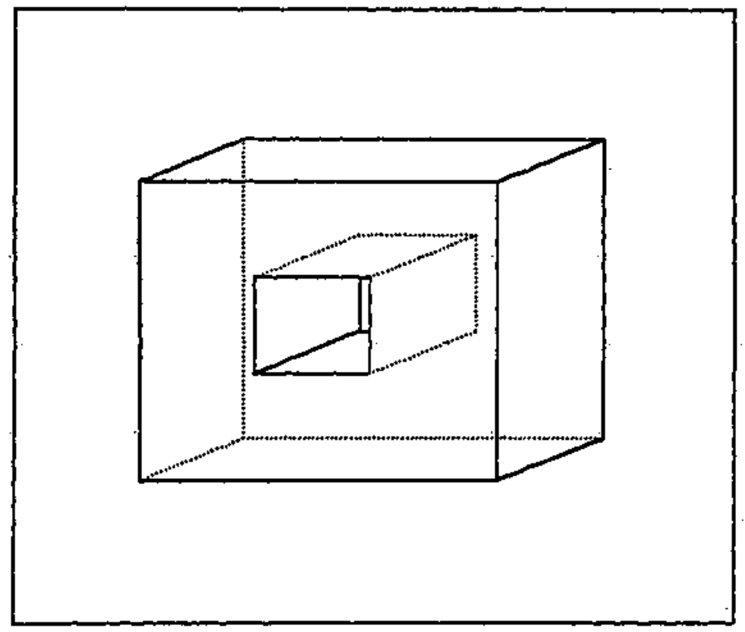


Figure 11
An Eulerian Picture Frame

A single proof is not enough to improve our original conjecture. We also need deeper proofs, as a counterweight against the narrowing of content entailed by lemma

[16] See [Poincaré 1893].

incorporation. Rule 4 is therefore added, in both its conservative and radical interpretation to the other three rules and the method is renamed *the method of proofs and refutations*.

There is one final amendment to the method. The existence of polyhedra which are not Eulerian suggests that we should try to find a *general* relationship between V, E and F which will apply to *all* polyhedra. This observation leads to a discussion of the method of *deductive guessing*[17] and to a sequence of generalizations of Euler's theorem. For example, for polyhedra with 'tunnels', we have:

(1) $$V - E + F = 2 - 2(n-1)$$

for an *n-spheroidal* polyhedron For a normal polyhedron, with no tunnels, $n = 1$, so (1) gives V – E + F = 2. For the picture-frame, which has one tunnel, $n = 2$, so (1) gives V – E + F = 0, which is correct. [ibid. p. 307-8]. For polyhedra with ring-shaped faces, like the crested-cube, we have:

(2) $$V - E + F = 2 - 2(n-1) + r$$

for a polyhedron with r ring-shaped faces. For the crested cube, we have a monospheroidal polyhedron ($n = 1$) and one ring-shaped face ($r = 1$). Hence (2) gives us: V – E + F = 2 – 2(1–1) + 1 = 3. The formula also explains the Eulerian picture-frame (Figure 11). This has one tunnel and two ring-shaped faces, so $n = 2$ and $r = 2$. Hence, by (2) the Euler number of this polyhedron is 2 – 2(2 – 1) + 2 = 2. [ibid. p. 309].

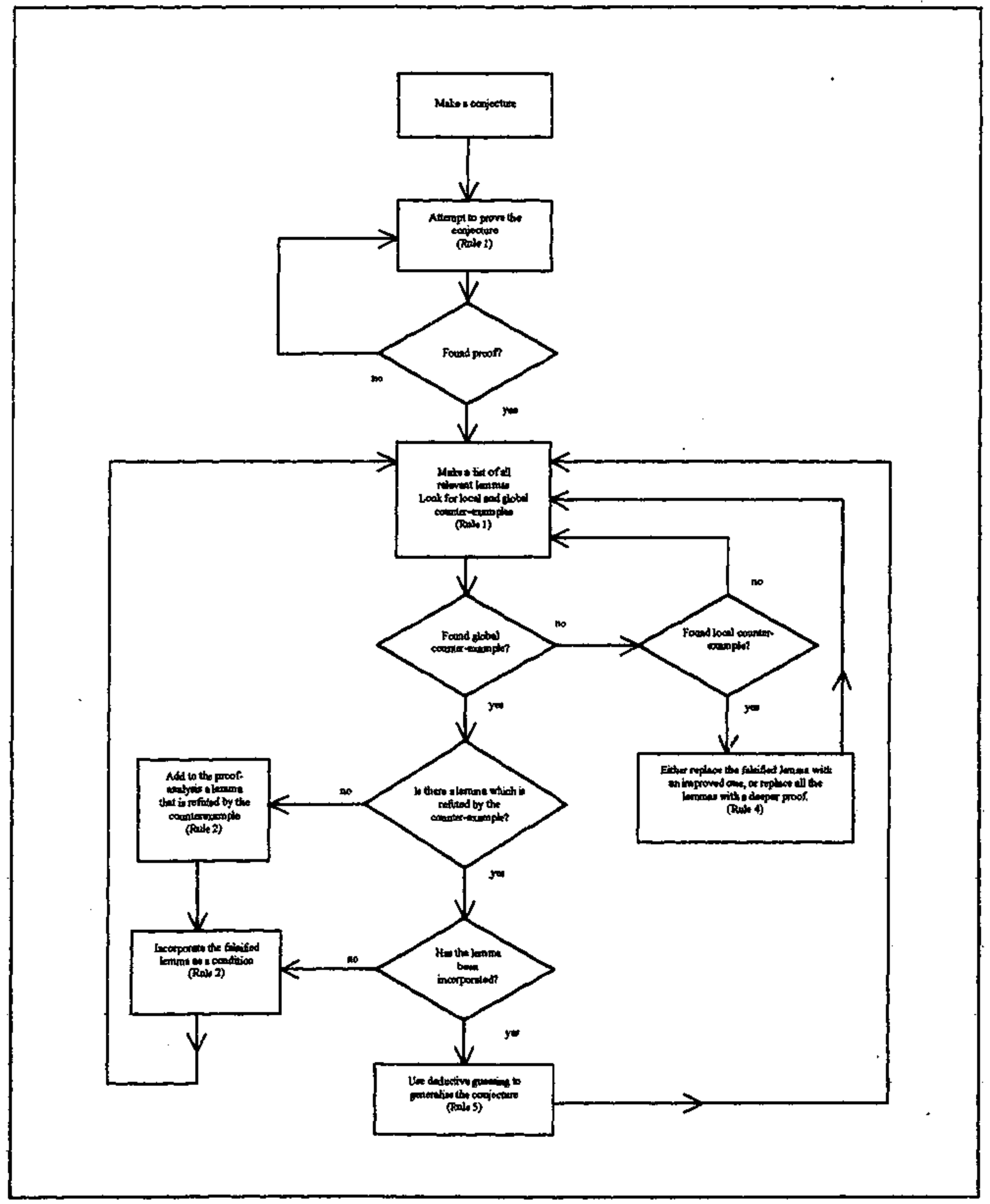
Local and global counter-examples then, can suggest that we need to generalise, by finding a deeper conjecture, one that covers more cases. A final rule is then added to the method of proofs and refutations:

[17] The idea here is to use a proof rather than induction to generate generalizations of a theorem. See [ibid. pp. 296-314].

> *Rule 5*: If you have counterexamples of any type, try to find … a deeper theorem to which they are not counterexamples any longer.
>
> [ibid. p. 306]

To summarise; the method of proofs and refutations counsels us, given a conjecture, to attempt to both prove and refute it. The method can only begin when we have *both* a proof and a counter-example. Indeed, proofs suggest counter-examples – by looking for counter-examples to the lemmas of the proof, we may find a counter-example to the main conjecture itself. If the counter-example is both global and local, we can use lemma-incorporation to improve the conjecture. If we find an apparently global, but non-local counter-example, we apply the principle of retransmission of falsity – we improve our proof-analysis by adding, if necessary, a suitable lemma to the proof-analysis, which will be refuted by the counter-example. On the other hand, we may find a local, but non-global counter-example. Such counter-examples are used to increase the content of our conjecture. We can either modify the refuted lemma slightly, or replace all the lemmas of the proof by finding a deeper proof, which makes use of lemmas which are not falsified by the counter-example. Finally, a global or local counter-example can also lead to an increase in content, by suggesting that we should look for a proof a more general conjecture, one that is not refuted by the counter-example, but which implies at as a special case. An overview of the method of proofs and refutations is shown in figure 12.

Figure 12
The Method of Proofs
and Refutations



3. The Significance of *Proofs and Refutations*

Disappointingly, there is very little critical analysis of Lakatos's work in the philosophical literature.[18] This is one reason why I have attempted to describe the method of proofs and refutations in some detail. Of course, my discussion has necessarily been schematic. I have skipped over large sections of the text which cover a wide variety of further topics. To do full justice to Lakatos's work would require a complete book all to itself.

The neglect of Lakatos's work by philosophers of mathematics may be due in part to the way in which *Proofs and Refutations* is written. As the reader may have noticed, the style of the work is a reflection of Lakatos's 'critical rationalist' view of the nature of inquiry. In the course of the dialogue, the students adopt positions on methodological questions which are continually subjected to criticism and debate. As a result, it is often hard to extract from Lakatos's work a clear and explicit statement of his final view. Furthermore, since the discussion centres around a single example, it is not obvious what sort of general account Lakatos is offering. Nonetheless, I think we can find in *Proofs and Refutations* many insights which are of crucial importance in the epistemology of mathematics and this is a further, more important reason for discussing Lakatos's work in some detail.

A common view of the epistemology of mathematics pictures the mathematician as beginning by laying down certain self-evident axioms and then proceeding to rigorously deduce theorems from them. In this way, a growing body of infallible knowledge is gradually built up. This is the picture suggested by the foundationalist account of the

[18] Two exceptions are [Kitcher 1977] and [Brown 1990].

structure of mathematical knowledge. Lakatos's central insight was that an examination of the actual history of mathematics will reveal the absolute poverty of this kind of picture of the growth of mathematical knowledge. If we look in detail at that history, we find something utterly at odds with the picture of a gradual accumulation of indubitably established truths, founded on a set of sure and certain first principles. What we actually see is something much more like the development of knowledge in the natural sciences.

In particular, the foundationalist account mistakes the role of proof in mathematics. Lakatos pointed out that a proof is only an *argument* and real mathematical proofs do not always begin from self-evident axioms, but are based on an often not explicitly formulated set of lemmas, which may be neither self-evident, nor indeed true at all. For this reason, the sort of justification a proof gives us need not be one that provides *conclusive* evidence that the theorem is true. By showing how a theorem depends on lemmas which are themselves open to rational doubt, a proof may in fact decrease our conviction that the theorem is true, rather than establishing it with certainty. By identifying counter-examples to the lemmas, one may be led to counter-examples to the theorem, or to ways of improving it. Thus, Lakatos argues, proofs are instruments of discovery, rather than instruments purely of justification. Proofs are one of the main tools in the development and growth of mathematical knowledge, but they are not the sort of perfect tool suggested by the foundationalist picture. They are a fallible tool in the development of a field of human knowledge which, for Lakatos, grows under the pressure of criticism and doubt, conjectures and refutations.

Another important fact that Lakatos's work reveals is that an account of the epistemology of mathematics need not depend on any particular metaphysical analysis of

the subject matter of mathematics. Lakatos proposes no such account and nothing he has to say about the epistemology of mathematics seems to depend on or require any particular ontology for mathematics. His analysis of the growth of mathematical knowledge is consistent with a wide variety of different ontologies for mathematics.[19] *Proofs and Refutations* is therefore, the first example of the sort of ontologically neutral, descriptive examination of the epistemology of mathematics which I argued for at the end of chapter two. This is another motivation for spending some time discussing the details of Lakatos's work here – it shows that this kind of investigation is not only possible, but also philosophically fruitful.

Of course, one can find fault with some of the details of Lakatos's account. The method of proofs and refutations applies most clearly to conjectures of the form 'All As are Bs'. Statements of this universal form can be refuted by a single counter-example, an object which is A, but not B. But clearly, not all mathematical statements have this form. What of mathematical statements that are either *existential* (such as 'There are numbers which cannot be expressed as a fraction') or *particular* (such as 'π is a transcendental number')? Such statements cannot be refuted by a single counter-example, so it is not clear how the method of proofs and refutations is meant to apply to them.

This is suggestive of a deeper worry. As already mentioned, Lakatos's argument is based on a single example, the history of Euler's conjecture. One might wonder then if

[19] James R. Brown makes this point in his article 'Proofs and Truth in Lakatos's Masterpiece' [Brown 1990]. It is worth pointing out however, that Lakatos's account may not be consistent with *all* accounts of the subject matter of mathematics. It is hard to see how *formalism*, for example, can accommodate the facts about the history and methodology of mathematics which Lakatos draws attention to. Nonetheless, any broadly *realist* account of mathematics would fit well with Lakatos's account. Ontology and epistemology are clearly not completely independent – a descriptive, ontologically neutral account of the epistemology of mathematics may rule out some kinds of account of the subject matter of mathematics and be at least suggestive of others.

Lakatos's account is more widely applicable, to other historical episodes and other branches of mathematics.[20]

Consider how Lakatos's account would apply to number theory. Lakatos argues that the growth of mathematical knowledge is driven by proofs and refutations of conjectures. The history of number theory, however is littered with conjectures that withstood all attempts to either prove or refute them for centuries – Goldbach's conjecture or Fermat's Last Theorem are obvious examples. The history of these conjectures does not, at first glance, appear to sit well with Lakatos's account.

However, looking more closely, we find that Lakatos can in fact account for these examples. Consider Fermat's Last Theorem. It is true that no one ever found a global counter-example to Fermat's conjecture – positive integers x, y and z that satisfy the equation $x^n = y^n + z^n$ for some $n \geq 3$. Nor was anyone able to find the proof which Fermat claimed to have discovered. But it would be wrong to say that proofs and refutations played no role in the development of our knowledge of Fermat's conjecture, which culminated in Wiles' final proof. Firstly, although unable to prove the conjecture itself, mathematicians were able to prove particular instances of it. Euler, for example, proved the case for $n = 3$. Although there are no *global* counter-examples to Euler's theorem, there are obvious *local*, but non-global counter-examples; that is, values of n greater than three for which the Fermat conjecture holds. Although such counter-examples do not of course *refute* the conjecture, they do play an important heuristic role in the method of proofs and refutations. Lakatos claims that they suggest we should look for a deeper proof and this is more or less

[20] In an appendix I of [Lakatos 1976], Lakatos does briefly consider another example – Cauchy's attempted proof of the claim that the sum of a convergent series of continuous functions is itself continuous. See also [Kitcher 1977; 1984, pp. 254-6]. We will come across Cauchy's proof again in section five of the current chapter.

what we find – a succession of increasingly general proofs which account for such local, non-global counter-examples, proofs for cases where $n = 4, 5, \ldots$ and so on and proofs such as Sophie Germaine's, which accounted for an infinite number of special values of n.[21] Hence, the history of Fermat's Last Theorem and of similar examples from number theory, can be made to accord fairly well with Lakatos's account, so long as we remember that it is not only global but also local counter-examples which play a role in the method of proofs and refutations.

This example illustrates an important fact however, which suggests a deeper criticism of Lakatos's account. I have argued that the case of $n = 4$ is a local, non-global counter-example to Euler's theorem that there are no positive integer solutions to the equation $x^n = y^n + z^n$ when $n = 3$. The claim that it is a local counter-example is trivially true, but we only know that it is a non-global counter-example by *proving* that the conjecture holds for $n = 4$.[22]

In some cases, showing that a particular object is a counter-example to a universal conjecture is a simple matter. We can draw polyhedra, for example and simply count the number of vertices, edges and faces. Likewise, we can check whether a particular triple of positive integers satisfies Fermat's equation for a particular value of n with a simple calculation. In general, if we have a universal statement 'All As are Bs' where B is an *effectively decidable* property,[23] then checking whether a particular object is a counter-example is straightforward. But of course, not all properties of mathematical objects are

[21] See chapter five, section five for further details and references.
[22] Notice also that the case of $n = 4$ is not more general than the case of $n = 3$. The theorem for $n = 4$, accounts for the local counter-example without also accounting for the case of $n = 3$. This is a state of affairs Lakatos does not consider – given a local, non-global counter-example the method of proofs and refutations suggests we should look for a more general proof. But this may not be possible – one can deal with the local counter-example with a theorem that applies only to that counter-example and not to others previously obtained.
[23] That is, one for which there exists an algorithm for deciding whether an object has the property or not.

effectively decidable. Where B is not a decidable property, it may be a very tricky matter indeed to find out whether an object is counter-example to the claim that 'All As are Bs'.

What this sort of example reveals is that refutation of a mathematical conjecture may require *proof.* This will, in general, also be the way in which conjectures which are not of universal form would be refuted. To refute the claim that a is F may require a non-trivial proof that a is not F. To refute the claim that some Fs are Gs, we need a proof that all Fs are not Gs. Likewise, showing that there are local, but non-global counter-examples to the theorem that for $n = 3$, there are no solutions to $x^n = y^n + z^n$, required a proof that there are no solutions for some values of $n > 3$.

This represents a major gap in Lakatos's account. For suppose we have a proof of a conjecture and a proof that some object is a counter-example to that conjecture. We need to decide whether to accept the refutation or reject it – since proof is never conclusive, either option is open to us. Lakatos's method of proofs and refutations does not tells us on what basis such decisions can be made.[24]

This problem is an example of the way in which the account offered in *Proofs and Refutations* inherits some of the difficulties which have been identified with Popper's account of the development of scientific theories[25]. Popper argued that no scientific law or theory can ever be conclusively confirmed and in this he was surely right. But he was wrong in thinking that scientific laws can be conclusively refuted – laws are much more resistant to falsification that Popper's simple logical model suggests. The reason for this, as many philosophers of science have since pointed out, is that scientific laws do not, by

[24] Kitcher makes just this point in relation to Lakatos's discussion of counter-examples to Cauchy's proof that a convergent series of continuous functions is continuous. See [Kitcher 1977, p. 783].
[25] It is worth pointing out that Lakatos himself was one of the philosophers responsible for pointing out some of these difficulties. See for example [Lakatos 1970, 1978].

themselves, entail any testable predictions at all. To derive a prediction from a law, we require *auxiliary hypotheses*. These may include not only singular statements of initial conditions, but also further laws. Hence, in a case where the prediction turns out to be false, we know that that either the law being tested, or one of the auxiliary hypotheses must be false, but logic alone cannot tell us which. In practice, the working scientist will not take an observation which is inconsistent with a law as definitively refuting it. The background assumption is that observational anomalies can and should be accounted for by rejecting or modifying one or more of the auxiliary hypotheses.[26]

This is analogous to the difficulty just mentioned with Lakatos's account of mathematical methodology. A refutation of a conjecture may be based on a proof which makes use of mathematical statements which can also be questioned. These latter statements are the analogue in mathematics of the auxiliary hypotheses required to test a scientific law. In such a situation, we know that either the original conjecture or one of the statements used to deduce the refutation must be false, but once again, we do not know which and Lakatos's account is silent on the criteria used by mathematicians in deciding the issue.

Lakatos's Popperian presuppositions are perhaps most clearly revealed in his discussion of induction in mathematics. Lakatos argues that the initial naïve conjecture is not arrived at by inductive generalization from an examination of particular cases:

> TEACHER: ...naïve guessing is not induction: there are no such things as inductive conjectures!
> BETA: But we found the naïve conjecture by induction! That is, it was suggested by observation, indicated by particular instances. . . . And among the particular cases that we have

[26] Many philosophers have made the point that no scientific law can be tested in isolation. See for example, [Duhem 1906, Quine 1951, Putnam 1974]. This point is discussed in more detail in chapter four.

> examined we could distinguish two groups: those which preceded the formulation of the conjecture and those which came afterwards. The former suggested the conjecture, the latter supported it....
> TEACHER: No! Facts do not suggest conjectures and do not support them either!
>
> [Lakatos 1963, p. 303]

The distinction made here, between induction as a method of discovering new truths and induction as a way of confirming or supporting conjectures is an important one. It may well be the case, as Popper argued, that inductive generalisation is not a good model of how scientific laws are discovered. A law may be discovered by a process of reasoning which is much more logically complex than simple induction (inference to the best explanation, analogy with the mathematical form of other laws and so on) as well as by entirely non-rational 'methods' – laws which are guessed at or suggested by dreams for example.

However, it does not follow from this observation that there is no such thing as inductive evidence for a conjecture. In chapter one I drew attention to the distinction between discovery and evidence. The means by which a truth is discovered may not be such as to provide us with any evidence for it, the evidence supporting the discovery may only be found at a later date. Since discovery and evidence are independent, the fact that induction is not a means of discovery does not entail that it cannot be a kind of evidence. Regardless of how a conjecture is first arrived at, confirmation of a wide variety of its instances can provide evidence that the conjecture is true. The history of mathematics reveals many examples of this kind of reasoning. Euler, for one, made extensive use of such arguments in his work. Of course, giving an account of this kind of evidence is notoriously problematic – I discuss some of these problems as they arise in the case of mathematics in chapter five. For now, it is enough to point out, that *pace* Popper and Lakatos, there is such a thing as inductive confirmation of a hypothesis or conjecture.

It is not just induction which Lakatos objects to however. He appears to want to do without the idea of any kind of confirming evidence at all. Here again, he is in agreement with Popper, who argued that the corroboration of a theory by verification of its observational consequences should not be taken as evidence that the theory is true[27]. Again, although Popper was right to say that we can never have *conclusive* evidence that a scientific theory is true, it is surely a mistake to infer from this that scientific theories are not confirmed to *any* degree by the evidence we have for them. We can have evidence for a scientific theory which although not conclusive, nonetheless gives us good reasons for thinking that theory is true.

Likewise, it is a mistake to think that one can do without an account of confirming evidence in the epistemology of mathematics, even if you agree that mathematics is a thoroughly fallible branch of inquiry. We have already seen how Lakatos's omission of such an account exposes a gap in the method of proofs and refutations; sometimes counter-examples to mathematical conjectures require sophisticated proofs and the question then arises how the principles on which such proofs depend are themselves justified. To answer this question, we need an account of the ways in which mathematical conjectures can be justified, as well as the ways in which they can be refuted. What Lakatos fails to provide is an account of mathematical evidence. Lacking such an account, he does not fully solve the problem of how a proof can provide us with mathematical knowledge. Lakatos was right to point out that proofs are instruments of discovery, but without an account of mathematical evidence, he is unable to show how they can also be instruments of justification.

---

[27] At least, this was Popper's position in his earliest writings. Later on, he introduced the concept of *verisimilitude*, which was intended to show how, through the process of corroboration and falsification, we arrive at theories which approach the truth more and more closely. However, his formal definitions of verisimilitude have been shown to face difficulties of their own. See [Miller 1974a, 1974b and Tichý 1974, 1978].

These problems however, should not distract us from the truth of Lakatos's central insight – that the history of mathematics provides us with a valuable testing ground for epistemological theories of mathematics. Although Lakatos's own theory may have its flaws, it certainly fairs much better in this regard than the alternative he was most concerned to attack, namely the foundationalist account.

The idea that the history of mathematics is not epistemologically irrelevant was revolutionary. Few philosophers of mathematics however have taken the point seriously. I would like now to examine the work of one notable exception, a philosopher who has built on Lakatos's insight and developed it in new and exciting ways.

4. KITCHER'S PROGRAMME

In *The Nature of Mathematical Knowledge*, Philip Kitcher develops a detailed and sophisticated account of the epistemology of mathematics which builds on many of the points which have emerged in our discussion of Lakatos's work. Like Lakatos, Kitcher rejects the thesis that mathematical knowledge is *a priori* or built on epistemologically secure foundations. He also emphasises the importance of the history of mathematics:

> A third break with the usual approaches to mathematical knowledge consists in my emphasis on the historical development of mathematics. I suggest that the knowledge of one generation of mathematicians is obtained by extending the knowledge of the previous generation. To understand the epistemological order of mathematics, one must understand the historical order....Most philosophers of mathematics have regarded the history of mathematics as epistemologically irrelevant. (Lakatos's principal insight was to recognize that this is a mistake).
>
> [Kitcher 1984, pp. 4-5]

Kitcher provides an *evolutionary* account of the epistemology of mathematics. The central idea of such an account is that at any time, the mathematical knowledge of a community is justified through its relationships to the knowledge of earlier communities. Kitcher summarises his account as follows:

> I shall explain the knowledge of individuals by tracing it to the knowledge of their communities. More exactly, I shall suppose that the knowledge of an individual is grounded in the knowledge of community authorities. The knowledge of the authorities of later communities is grounded in the knowledge of the authorities of earlier communities. Putting these points together we can envisage the mathematical knowledge of someone at the present day to be explained by reference to a chain of prior knowers. ... However, if this explanation is to be ultimately satisfactory, we must understand how the chain of knowers is itself initiated. Here I appeal to ordinary perception. Mathematical knowledge arises from rudimentary knowledge acquired by perception.
>
> [ibid. p. 5]

Kitcher's suggestion is that mathematics begins as perceptually justified. This gives us some basic knowledge of arithmetic and geometry. Mathematicians then inherit the mathematical knowledge of the previous generation, modify and extend it in identifiably rational ways and pass it on to the next generation. At each stage, these modifications and extensions (proofs of new theorems, introduction of new language, concepts and so on) are justified via their relationship to the prior body of knowledge which mathematicians have inherited. In this way, our mathematical knowledge evolves from a primitive, perceptually justified body of knowledge concerning simple facts of arithmetic and geometry, into the wide-ranging, sophisticated and abstract system of contemporary mathematics.

Hence, for Kitcher, the epistemological order of mathematics is revealed in its historical development. This does not mean that the historical order and the epistemological order correspond exactly. The grounds on which some part of mathematics is initially accepted might fail to properly justify it; that justification may only be given much later. The point of Kitcher's evolutionary account is just that this later justification will proceed by showing how that part of mathematics is appropriately related to prior mathematical knowledge. The historical and epistemological order can diverge at times, but '[w]hat matters is that we should be able to describe a sequence of transitions leading from perceptually justified mathematical knowledge to current mathematics. In giving this description .... we shall appeal to antecedently justified principles to justify further extensions.' [ibid. p. 9].

Kitcher introduces the theory of the *ideal collector*, mentioned in the previous chapter, in order to explain how the mathematics passed on from one generation to the next can have its origin in a perceptually justified body of knowledge. The theory is meant to provide a non-platonist, constructivist account of the ontology of mathematics. The basic idea is that we should think of mathematics as an idealised theory of what human beings can *do*. In particular, it is an idealised theory of the operations of counting, collecting and ordering objects. Kitcher argues that similar constructivist accounts such as intuitionism fall into the trap of attempting to delineate in advance what operations the 'ideal subject' can perform and thereby severely restrict the amount of mathematical knowledge we can be said to have. Kitcher places no such restrictions on ideal operations and in this way hopes to account for the entirety of classical mathematics [ibid. pp. 101-38].

Some basic knowledge of what operations the ideal subject can perform can be gained through ordinary perception; 'we observe ourselves and others performing particular acts of collection, correlation and so forth, and thereby come to know that such operations exist. This provides us with rudimentary knowledge – proto-mathematical knowledge, if

you like.' [ibid. p. 117]. This 'proto-knowledge' is turned into knowledge of arithmetic, geometry, set-theory and so on, by a process of generalising, systematising and idealising the operations we have seen ourselves and others carrying out.

Like Lakatos, Kitcher sees the development of mathematics as proceeding in ways which are analogous to the development of the sciences. But Kitcher is able to improve on Lakatos's model of the development of mathematics by taking into account some of the problems with the Popperian model of scientific methodology mentioned in chapter three.

Recall how Popper's model is open to the objection that it overestimates the extent to which theories can be falsified by observation. Discrepancies with observation are often accounted for by modification of auxiliary hypotheses; the falsification of a hypothesis by observation does not necessarily lead to its outright rejection. Thomas Kuhn has famously argued that not only are scientific theories resistant to falsification in this way, but that they are falsified almost all the time. Theories are constantly in conflict with both observation and other parts of scientific theory. The attempts of scientists to account for such anomalies makes up a large part of what Kuhn calls *normal science*. When all attempts to account for a theory's external and internal problems start to fail, scientific practice may enter a revolutionary period and only then will the theory be finally abandoned and replaced with a more promising alternative.[28]

Kitcher argues that what this shows is that observation and experiment are not the only source of scientific change. A scientific theory may also come into conflict with other theoretical principles. These internal problems are of various kinds. Newton worried about instantaneous action at a distance, for example because it seemed contrary to the notion of a mechanical action. Contemporary physicists attempt to create unified theories, connecting

[28] See [Kuhn 1970].

the fundamental forces of nature - here we find tensions for example, between the principles of quantum mechanics and those of general relativity. There can also be internal tensions in a theory. For example, in quantum mechanics, some have argued that there is a conflict between the deterministic component of the theory (the evolution of the wave equation) and the non-deterministic component (which describes what happens when a measurement is made).[29] In cases like these, we see science evolving in ways that are not prompted by purely observational evidence, but by inter- and intra-theoretical problems. Kitcher argues that this kind of evolution is the rule in mathematics:

> There are always "internal stresses" in scientific theory, and these provide a spur to modification of the corpus of beliefs. I propose to think of mathematical change as akin to this latter type of modification. Just as the natural scientist struggles to resolve the puzzles generated by the current set of theoretical beliefs, so too mathematical changes are motivated by analogous conflicts, tensions, and mismatches.
>
> [Kitcher 1984, p. 154]

One example of the kind of theoretical puzzle that Kitcher sees as driving the development of mathematics is the conflict between early methods of differentiating functions and the normal rules of algebra and arithmetic. The resolution of these and other theoretical puzzles led to the eventual rigorization of the calculus by mathematicians such as Cauchy and Weierstrass – a story which Kitcher tells in detail in the final chapter of his book.

Kitcher extracts from Kuhn's work the idea that we should see the development of science as consisting of the modification and extension of scientific practices. This development is sensitive to changes in the results of observation and experiment, but it is

[29] See for example [Penrose 1989, pp. 250-1].

also sensitive to changes in other components of the practice; language, theoretical principles, problems, approved methods of reasoning, methodological maxims and so forth.

Hence Kitcher introduces the concept of a *mathematical practice*:

> One of Kuhn's major insights about scientific change is to view the history of a scientific field as a sequence of practices: I propose to adopt an analogous thesis about mathematical change. I suggest that we focus on the development of mathematical practice, and that we view a mathematical practice as consisting of five components: a language, a set of accepted statements, a set of accepted reasonings, a set of questions selected as important, and a set of metamathematical views (including standards for proof and definition and claims about the scope and structure of mathematics).
>
> [ibid. p. 163]

Given this Kuhnian framework,[30] Kitcher's account aims at showing how mathematical knowledge grows through the rational modification of mathematical practices; '[t]he problem of accounting for the growth of mathematical knowledge becomes that of understanding what makes a transition from a practice ... to an immediately succeeding practice a rational transition' [ibid. p. 164].

5. The Evolution of Mathematical Practices

Let us look at some of the components of a mathematical practice in more detail. The set of *accepted statements* of a mathematical practice is the set of sentences, formulated in the language of the practice, which are accepted as true by the mathematical community of the time. The ways in which new statements can be added to the set of accepted statements are fairly obvious, a statement can be added when it is proved for the

[30] Although he adapts Kuhn's idea of a scientific paradigm to yield the notion of a scientific (or mathematical) practice, Kitcher rejects some of Kuhn's additional claims about scientific change. In particular he rejects the thesis that pre- and post-revolutionary paradigms are incommensurable. [ibid. pp. 162-3].

first time, or perhaps accepted on grounds which fall short of proof, but which are sufficient according to the standards of the day.

But mathematics is not a simple cumulative process of adding more and more statements to our store of truths. Statements that are accepted at one time, may later come to be rejected. This might happen when a flaw is found in a proof of an accepted theorem, or when we find a counter-example to a statement which was accepted on the basis of inductive evidence. But removal of a statement need not involve an outright rejection of it. Early sixteenth century mathematicians would have accepted the statement 'there is no number whose square is negative'. Contemporary mathematicians would reject this statement. But this hides a certain measure of agreement; some of the content of the sixteenth century statement is preserved in the currently accepted statement that 'there is no *real* number whose square is negative.' [ibid. p. 179].

The set of *questions* of a mathematical practice is Kitcher's representation of the set of *problems* which occupy the mathematical community of the time. A question is usually removed from the practice by being answered – this represents a problem being solved. On the other hand, questions may be removed, not because they have been answered, but because they are no longer considered important. Changes in the perceived importance of a field of mathematics, or changes in the requirements of applications may result in a question being removed from the set of questions considered worth asking by the practice. Of course, a question removed for these reasons may later return to the practice if its solution comes to be seen as important in a new way. [ibid. p. 185-7].

The set of *accepted reasonings* of a mathematical practice is the set of all those sequences of statements which mathematicians of the time advance as arguments in support

of members of the set of accepted statements. Kitcher makes several important distinctions among the members of the set of accepted reasonings. Firstly, they can be divided into classes according to the degree of support they provide for their conclusions. In this respect, an important class of accepted reasonings are of course *proofs*. Proofs are those arguments which provide *optimal support* for their conclusions (relative to the standards of proof accepted at the time – these standards forming part of the metamathematical component of the practice) but they also have an *explanatory* function – by revealing the logical connections between their conclusions and other accepted statements of the practice, proofs improve our understanding of a mathematical field. [ibid. p. 181].

A second class of reasonings are the various *non-deductive* arguments; arguments by induction or analogy for instance. Such arguments provide the lowest degree of support to their conclusions. A third, intermediate type, are what Kitcher calls the *unrigorous reasonings* of the practice. These are arguments which cannot be reformulated as valid deductions in their accepted system of proofs. Seventeenth and eighteenth century calculus was full of such reasonings. Kitcher makes use of the following example [ibid. p. 182]. Mathematicians at this time would calculate the derivative of the function $y = x^2$ along the following lines. Let x increase by an infinitesimal amount Δ. Then:

$$\begin{aligned}\frac{dy}{dx} &= \frac{\left((x+\Delta)^2 - x^2\right)}{\Delta} \\ &= \frac{\left(x^2 + 2x\Delta + \Delta^2 - x^2\right)}{\Delta} \\ &= \frac{2x\Delta + \Delta^2}{\Delta} \\ &= 2x + \Delta\end{aligned}$$

But since Δ is infinitesimally small in comparison to x, we can ignore it. Thus:

$$\frac{dy}{dx} = 2x$$

This looks enough like an ordinary algebraic demonstration to suggest that it could be reformulated as a rigorous proof. It is not sound as it stands however. We first divide through by Δ and then take Δ as zero. But division by zero is not normally allowed. Why can we get away with it here? Although unrigorous reasonings like these often give us the right answers, they reveal a certain sort of failure of understanding. We can use the techniques they offer to get solutions to problems, but we do not really understand why the techniques work. Kitcher argues that unrigorous reasonings have been a major source of intra-theoretic tension in mathematics, the resolution of which has played an important role in the evolution of mathematical knowledge. [ibid. pp.182, 213-217].

The members of the set of accepted reasonings can also be divided into classes according to the ways in which they can be used to answer questions. In this context, Kitcher makes a distinction between those arguments which provide *problem-solutions* and those which do not. An argument is a problem-solution if it enable us to obtain an answer to a question even if we do not already know the answer. A *confirmation-technique*[31], by contrast, enables us to confirm that an answer to a question is correct, but will not generate the solution for us if we have not already conjectured an answer.

Proof by mathematical induction is an example of a confirmation-technique. We can use it to confirm for example, that $1+2+3+4+...+n=\frac{1}{2}n(n+1)$, but mathematical induction will not generate an answer to the question 'what is the sum of the first n natural

[31] I have introduced the term *confirmation-technique* here to stand for any argument which is not a problem-solution.

numbers?' for us[32]. We can only use it to check answers once we have already got them by some other means. Hence not all proofs are problem-solutions. But neither are all problem-solutions proofs. An example of a problem-solution which is not a proof is the unrigorous reasoning described above which can be used for generating answers to questions about the derivatives of polynomial functions. Another example which Kitcher discusses is Euler's technique for finding sums of infinite series. Euler used this technique to discover, for example that $1+\frac{1}{4}+\frac{1}{9}+\frac{1}{25}+...=\frac{\pi^2}{6}$. Although it is not a proof, Euler's technique does enable us to generate answers to questions we do not already know the answers to, and hence counts as a problem-solution.[33].

On the other hand, although some confirmation-techniques are proofs (mathematical induction for example) some are not. An example of a confirmation technique which is not a proof would be the technique Euler used for checking the results of his method for summing infinite series; he computed the sums of a finite number of terms of the series and verified that they approximated to the values his technique produced. Another example would be confirming a universally quantified statement by checking that the theorem holds for a few particular cases, as when we check that Euler's formula V-E+F=2 holds for the regular polyhedra. Kitcher calls confirmation-techniques which are not proofs *checking-procedures*. [ibid. p. 182-3].

In the penultimate chapter of his book, Kitcher identifies and discusses five patterns of change in mathematical practice, changes which involve modification of several

[32] See chapter five, section four.
[33] Euler's argument is discussed in more detail in chapter five. See also [Kitcher 1984, pp. 196-7] and [Polya 1954a, pp. 17-21].

components of the practice at once. These are *systematization, generalization, question-generation, question-answering* and *rigorization*.[34]

Systematization is a modification of mathematical practice with the aim of improving our understanding of previous material by achieving a systematic presentation of previous results. This may involve the introduction of new statements or language into the practice. For example, Cayley gave a definition of the concept an abstract group with the aim of systematizing a great deal of prior work in abstract algebra and Lagrange introduced new concepts which provided a systematic treatment of techniques for solving equations. [ibid. pp. 217-224].

Generalization is a modification of mathematical practice which shows how previous results can be seen as special cases of a more general theory. New language, statements and reasonings are introduced in such a way that we obtain analogs of some previous results and are able to show how others can be seen as special cases of the more general perspective. Kitcher central example is Cantor's introduction of the theory of transfinite numbers as a generalization of finite arithmetic [ibid. pp. 207-12].[35]

Modifications to mathematical practice such as these may provoke new questions – this is the pattern of *question-generation.* As expressions for new kinds of object are introduced into the practice, new questions may arise by analogy with old ones. [ibid. pp.187-188]. Additions to the set of accepted statements may also generate new questions. Suppose for example, that we come to accept that some objects have a certain property, while others do not. Then the question naturally arises whether we can find a condition

[34] In his discussion of changes in the language component of a mathematical practice, Kitcher also identifies a sixth pattern of change, which might be called *reinterpretation.* See [Kitcher 1984, pp.158-61]. See also section five, below.
[35] I discuss this example in more detail in chapter five, section two.

which will separate those objects which have the property from those that do not. Kitcher gives an example from the theory of equations; '[a]fter Abel's discovery that there is no method for solving the quintic in radicals and Gauss's discovery that some classes of equations of high degree can be solved in radicals, Galois posed the (immensely fruitful) question "Under what conditions will there be a method for solving an equation in radicals?"' [ibid. pp. 188, 204-5].

Question-answering is a modification of mathematical practice whereby new statements and reasonings are introduced in order to answer one or more of the accepted questions of the practice. This may also involve modification of the language component of the practice; the introduction of terms for new entities, for example, or new language for describing objects already discussed.

The new statements, reasoning and language introduced into the practice in this way may bring in their wake new problems and questions. The new language may not initially be well understood, the reasonings may not be rigorous, or there may be apparent counter-examples to some of the new statements. However, Kitcher argues that this kind of modification to a mathematical practice is justified, or rational, to the extent that the new techniques '..enable the mathematical community to answer questions previously recognized as important, and the value of providing answers .. outweigh the difficulties involved in the ill-understood language, the new statements and the unrigorous reasonings.' [ibid. p. 195].

A further condition on the rationality of such a change is obviously that there must be reasons for believing that the proposed answers to the questions are *correct.* Otherwise, there would no reason why we should not adopt the statement 'every even number greater than two is the sum of two primes' merely on the grounds that it answers the question 'Is Goldbach's conjecture true?'. If the modification is not to be *ad hoc*, there must be independent grounds for accepting the answers, grounds other than those provided by the new reasonings and statements.

However, it may be that although the antecedent practice contains some way of checking that proposed answers to questions are correct, the new modification provides a way in which answers can be discovered. That is, the antecedent practice may contain no *problem-solution* reasonings for the questions, although it does contain *confirmation-techniques* for answers, once they have been obtained. The confirmation-techniques of the prior practice may not even include any proof techniques; for example, the practice may only contain techniques for getting approximate answers to the questions, whereas the new modification allows us to derive exact answers. On the other hand, the extension might provide a general solution to a question, where the previous practice only provides techniques for solving some particular instances or special cases. Even if the antecedent practice contains methods for generating answers to *all* the questions, the new modification may provide a systematic way of answering those questions, where the old methods were a disparate, unsystematic 'bag of tricks'. [ibid. pp.195-6].

One example of this kind of pattern of change which Kitcher discusses is Descartes' creation of analytic geometry. By introducing the idea of a co-ordinate system and showing how geometric curves could then be represented by equations, Descartes' theory synthesized geometry and algebra, thereby allowing algebraic techniques to be applied to geometric problems. Descartes was able to use his new approach to answer questions in geometry which could not be solved using previous techniques – questions whose solution

had baffled the ancient geometers for example. Nonetheless, those previous techniques could be used to check some of the answers which Descartes obtained using his new method. In particular, he was able to give general solutions to problems concerning the construction of loci, which could be verified by checking them against the partial solutions for the special cases which the Greeks had managed to obtain. [ibid. pp. 197-8].

The final pattern of change is *rigorization*. Recall that the *unrigorous reasonings* of a mathematical practice are those arguments which although successful, cannot be reconstructed as arguments which conform to the standards of rigorous proof required by the practice at the time. A modification to mathematical practice which introduces rigorous replacements for certain unrigorous reasonings of the practice is justified by showing how it explains previously accepted results; our understanding of those results is improved by showing how they can be rigorously deduced from certain statements in a way that explains the success (and failures) of the unrigorous reasonings.

As in the case of question-answering, it may happen that the proposed modification of mathematical practice brings with it new problems and difficulties, which need to be weighed against the benefits that rigorization brings. Kitcher argues that the rigorization may nonetheless be rationally acceptable if the costs are outweighed by the benefits, where the benefits include not only the provision of rigorous replacements for unrigorous reasonings, but also the ability of the proposal to further research on important questions. [ibid. p. 217].

In Kitcher's view a *necessary* condition for it to be rational to demand a rigorization of a mathematical practice is obviously that the practice must contain some unrigorous reasonings. However, he denies that this is a *sufficient* condition. Consider

Cauchy's proposals for reconstructing unrigorous arguments in analysis by means of the concept of a *limit*. Why were his proposals almost immediately adopted by the mathematical community? It cannot be merely because they provided rigorous replacements for unrigorous arguments. The fact that arguments in analysis were unrigorous had been recognized since the invention of the calculus in the early seventeenth century and the idea of using the limit concept was not new – it is prefigured in some writings of Newton himself for example. Why then, were the problems of rigour more or less ignored until the early nineteenth century? Why did they suddenly become so important that Cauchy's reconstruction of analysis was required? In order to better understand Kitcher's answers to these questions, it will be helpful to summarise some of his discussion of the historical development of analysis. The main points of Kitcher's argument will be most clearly illuminated by the looking at the history of research on infinite series.

Leibniz was one of the first mathematicians to show how the new techniques of the calculus could be used to answer questions about the sums of infinite series. For example,[36] having shown that $\int_0^1 \frac{1}{1+x^2} = \frac{\pi}{4}$, Leibniz expands the function $\frac{1}{1+x^2}$ as a power series, obtaining:

(1) $$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \ldots.$$

Leibniz then integrates this series term by term and obtains:

(2) $$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots..$$

[36] For more details of this and similar examples see [Kitcher 1984, p. 242] and [Kline 1972, pp. 436-66].

Euler was later to develop Leibniz's method into a powerful set of techniques for finding the sums of infinite series.[37] However, the technique of representing functions as a power series was known to occasionally yield anomalous results. For instance, using the same method, Leibniz obtained:

(3) $$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots...$$

Setting $x = 1$ in this expansion we get:

(4) $$\frac{1}{2} = 1 - 1 + 1 - 1 + 1 - 1 + \ldots..$$

Leibniz argued that we can explain away this apparently anomalous statement. If we take any finite *even* number of terms (e.g: $1-1, 1-1+1-1$ and so on) the sum is equal to zero. If we take any finite *odd* number of terms (e.g $1, 1-1+1, 1-1+1-1+1$ and so on) the sum is equal to 1. Hence, according to Leibniz, "...it follows that when we proceed to the case of an infinite number of terms, where the even and odd cases are mixed, and there is equal reason for it to go either way, we obtain $\frac{0+1}{2} = \frac{1}{2}$."[38] – that is, the infinite sum is the average of the finite sums for odd and even numbers of terms.

Euler criticized Leibniz's explanation, arguing that only the sum of a *convergent* series is approximated by the sum of a finite number of its terms and the expansion $\frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots..$ only converges when $x < 1$[39]. Nonetheless, Euler refused to

[37] See [Kline 1972, pp. 446-54].
[38] Cited in [Kitcher 1984, p. 242].
[39] Euler also pointed out that if we set $x = 2$ in the expansion $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots.$ we get the even stranger result that $-1 = 1 + 2 + 4 + 8 + \ldots$. The sums of finite numbers of terms of this series get bigger and bigger without limit, so we cannot explain away the result along the lines suggested by Leibniz's argument

say that divergent series do not have a sum. If by means of a series expansion and appropriate substitutions, we can obtain identities like that of (4), then the value obtained in this way can be thought of as the 'sum' of the divergent series. Euler made frequent use of this concept of the sum of a divergent series to obtain his results on the sums of convergent series – results which *could* be verified using the technique of computing partial sums. Kitcher argues that the verifiable success of Euler's methods for finding sums of convergent infinite series, made it rational for him to use those methods, despite the lack of rigor in the arguments. [ibid. p. 243-4].

These problems of rigor however, assumed a new urgency in the nineteenth century as certain questions about infinite series came to the forefront of mathematical research. At the very beginning of the nineteenth century, Joseph Fourier showed how trigonometric series expansions of functions could be used to solve important problems in physics. A trigonometric series has the general form:

$$\frac{1}{2}a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + (a_3 \cos 3x + b_3 \sin 3x) + \ldots\ldots$$

where a_n and b_n are constants. Fourier was able to show that a wide class of functions can be represented as a sum of sines and cosines in this way and he was able to derive a formula for the coefficients a_n and b_n of the trigonometric series which represents such a function. He then applied these results to problems concerning heat flow[40]. The use of trigonometric series quickly become an important technique for solving partial differential equations which arose in other areas of physics.

for the case of (4). That argument invokes considerations about the sums of finite segments of the series, but in Euler's view, *divergent* series do not have sums in a sense which makes such considerations appropriate.

[40] In his *Analytic Theory of Heat* [Fourier 1822]. See also [Kline 1972, pp. 966-72] and [Stewart 1987, pp. 228-31].

This work immediately generated an important question. Fourier had shown that *some* functions can be expressed as trigonometric series – can *every* function be expressed as a trigonometric series? Kitcher calls this the *Fourier question*. The Fourier question is just one example of a general trend; questions about series representations of functions were to become increasingly important in the mathematics of the nineteenth century. Little progress was made on these questions however because the basic concepts of continuity, convergence and series sum were not well understood. The lack of rigour in the theory of infinite series was now becoming a serious obstacle to the solution of the mathematical problems considered most important by the mathematical community.

In his *Cours d'Analyse* [1821] Cauchy proposed that the concepts of continuity, convergence, series sum and derivative should be defined in terms of the algebraic concept of a *limit*. In particular, he introduced the following definition of convergence and sum of an infinite series:[41]

(D1) An infinite series $\sum_{n=0}^{\infty} s_n$ is *convergent* if and only if the sequence of partial sums $\sum_{n=0}^{N} s_n$ tends to a limit as N tends to infinity and this limit is the *sum* of the series.

Notice that in (D1), Cauchy explicitly rejects the idea that a divergent series can have a sum. Hence Cauchy's proposal involved the rejection of Euler's techniques for solving problems about the sums of infinite series of numbers. By Cauchy's time however, questions about the sums of infinite series of *numbers* were seen as less important than

[41] See [Kitcher 1984, p. 246] for this and Cauchy's definitions of continuity and the derivative.

questions (such as the Fourier question) concerning the representation of arbitrary functions by an infinite series of *functions*. Since Cauchy's definitions provided tools for tackling the new questions, it was justifiably adopted, even though it eliminated some successful reasonings of the prior practice. [ibid. p. 250].

However, Cauchy's rigorization of the calculus was known to give rise to new difficulties. Consider the Fourier question. Euler had argued that it is unlikely that every function can be represented as the sum of a trigonometric series. The trigonometric functions have special properties - they are both continuous and periodic - properties which do not apply to the vast majority of functions. How could a discontinuous function, for example, be represented by a sum of a continuous functions? [42]

In the *Cours d'Analyse*, Cauchy attempted to give a rigorous version of this informal argument by proving that *the sum of an infinite series of continuous functions is always continuous*. If this is right, then the Fourier question would be settled in the negative; not every function can be represented by a Fourier series, since no discontinuous function can be represented by a sum of continuous functions. Unfortunately, Cauchy's attempt at a proof was a failure. Abel provided a counter-example to Cauchy's theorem in 1826. The series:

$$\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \frac{1}{4}\sin 4x......$$

is a convergent series of continuous functions, but it converges to a function which is *discontinuous* at each value of $x = (2n+1)\pi$. [ibid. p. 254, see also Kline 1972, p. 965]. The problem with Cauchy's argument, from a modern perspective, is his failure to

[42] Euler first proposed this argument in his debate with D'Alembert on the problem of the vibrating string. See [Kitcher 1984, p.245,249].

distinguish between two different ways in which an infinite series can converge to a function; what we would now call *pointwise* convergence and *uniform* convergence. It was Weierstrass who was to most clearly demonstrate where Cauchy had gone wrong, by introducing the modern ε-δ formulation of the concepts of continuity and convergence in which the distinction can be clearly formulated. [ibid. p. 254-9].

Kitcher compares the foundational work in analysis to Frege's investigations into the foundations of arithmetic and argues that his account of rigorization can explain why Frege's work was more or less ignored by the mathematical community. If we think of Frege's programme as a proposal for the rigorization of reasoning in elementary number theory, the explanation is not hard to find. The fact that reasoning in number theory was, to some extent unrigorous, was not by itself sufficient for it to be rational to adopt Frege's formal definitions and proofs as rigorous replacements for those reasonings. The lack of rigor in this area posed no real threat to mathematical research. [ibid. p. 268].

Michael Dummett in *Frege Philosophy of Mathematics*, argues that Kitcher is selling Frege short here. In his discussion of the sections of the *Grundlagen* where Frege sets out the motivation for his work, Dummett writes:

> There would be little point in dwelling on the reasons Frege gives for undertaking the investigation, so obvious must its interest appear to almost all with any philosophical inclination, were it not that there has been a recent movement, led by Philip Kitcher, to argue that it was indeed pointless. The argument is that, unlike the clarification of the foundations of analysis, it was not needed for the resolution of antinomies hampering the progress of mathematics. This might be thought the expression of a philistine attitude towards philosophy on the part of certain mathematicians by anyone unaware that it actually proceeded from philosophers....The questions what the natural numbers are, and how we know what we assume to be true about them, are of intrinsic interest, whether or not the answers contribute to progress within number theory....they lack interest only if either number theory itself is of no value, or philosophy as a whole is devoid of interest.
>
> [Dummett 1991, p. 11]

I hope it is clear from my discussion of Kitcher's account of rigorization why Dummett's criticism is misdirected. Kitcher does not say that Frege's investigation was *pointless*. Nor does his argument have anything to do with the status or interest of *philosophical* questions about mathematics. I am sure that Kitcher would not deny that Frege's work represents an important contribution to philosophy. The point is not that philosophical investigations into mathematics are only useful if they contribute to an increase in our mathematical knowledge. Rather, the point is that a proposal to *replace* an unrigorous field of mathematics with a rigorous reconstruction of it, cannot be justified merely on the grounds that the field in question is unrigorous. It is perfectly reasonable for mathematicians to use techniques and methods which are imperfectly rigorous, if those techniques enable them to successfully solve problems that interest them.

Mathematicians, like most scientists, are pragmatists. They are not moved by purely philosophical worries about the clarity of concepts or rigor of arguments. Just as most mathematicians did not let Berkley's criticisms of the calculus worry them too much, since the techniques of the calculus obviously worked, so mathematicians did not worry too much about giving a rigorous definition of the concept of number – number theory was, for the moment, getting on just fine without any such definition.

Of course, this does not mean that Frege's work was of no mathematical interest. As Dummett points out, '[p]lainly, inventing modern mathematical logic, and devising the very first formal system, were major contributions to mathematics under any but the narrowest circumscription of what constitutes mathematics' [Dummett 1991, p. 12]. Frege's work introduced many important mathematical ideas – the concept of the ancestral of a relation and of definition by logical abstraction, for example. In addition, we should

not forget, that despite the contradiction in Frege's system, an important and interesting result – Frege's Theorem - can be salvaged from it. But all of this is entirely consistent with what Kitcher says. Although Frege's work was certainly an important contribution to pure mathematics (let alone philosophy) it is nonetheless true that there is no great mathematical pay-off, in terms of problem-solving power, to be gained by adopting his demand to replace the unrigorous reasonings and definitions of the practice with his formal methods and this explains why his suggestions were not immediately and universally adopted, in the way that Cauchy's proposals were.

***

The patterns of change which Kitcher discusses are not mutually exclusive. We have seen how considerations of problem-solving power – justification by question-answering – may be involved in the process of rigorization. Likewise, a modification to practice may be justified not only on the grounds that it provides a generalization of previous material, but because it also enables us to answer questions – the introduction of complex numbers would be an example. A generalization may also provide a unified way of deriving previous results, and so be further justified on the grounds of its ability to systematize those results. The great power of Kitcher's account lies in its ability to explain the evolution of mathematics by showing how the different kinds of mathematical changes which he identifies can work in combination to produce large-scale changes to mathematical practice.

6. ORIGINS, KNOWLEDGE AND JUSTIFICATION

The Nature of Mathematical Knowledge represents one of the most important contributions to the epistemology of mathematics in recent years. Kitcher develops the idea, implicit in the work of Lakatos, that an adequate epistemology for mathematics should be faithful to mathematical practice with a high degree of philosophical rigour and attention to historical detail. The thesis that mathematics is a science is also advanced by showing how the development of mathematics can be explained by making use of a more sophisticated model of the evolution of science. Notice that Kitcher's evolutionary account of the development of mathematics through rational inter-practice transitions could in general, apply equally well to the development of physics, or indeed any other field of scientific inquiry.

It is worth noticing that Kitcher's account of the subject matter of mathematics, the theory of the ideal collector, plays hardly any role at all in his account of the *evolution* of mathematics. What he has to say about the justification patterns of change in mathematics such as generalisation, systematisation and question-answering is quite consistent with any number of alternative accounts of the ontology of mathematics, including platonism.

Of course, the theory of the ideal collector is really meant to account for the *origins* of mathematics, rather than its evolution, by showing how the chain of mathematical practices can have a perceptually warranted beginning. It does this by providing an account of the subject matter of mathematics which allows for perceptually acquired knowledge of that subject matter. But even here, what Kitcher has to say in general about the perceptual origins of mathematics is consistent with many alternatives to his preferred ontology. As we shall see, Quine has argued that there can be empirical evidence for mathematics and

hence that at least some mathematics can be perceptually justified. The details of Quine's argument and the question of whether or not it is a success will occupy us in the next chapter. The point I want to make here is that Quine's argument is quite independent of any specific ontology for mathematics. In particular it is quite consistent with platonism and indeed Quine himself takes his argument as showing how we can have perceptually grounded knowledge of abstract objects. Of course, Kitcher thinks that platonism faces other problems, quite apart from the problem of showing how we can have knowledge of abstract objects; problems which his theory of the ideal collector is designed to avoid.[43] However, my present point is just that the ideal collector is not necessarily required in order to show how there could be a perceptual grounding for the chain of mathematical practices, since there may be ways of showing how there can be perceptual evidence for mathematics which do not require any specific ontology for mathematics.

As Kitcher points out, an evolutionary account of mathematical knowledge must explain both the origins of mathematical knowledge as well as giving an account of the growth of mathematical knowledge. In the same way an evolutionary account of life should explain both how life got started and how it then evolves. Notice however, that in the case of the theory of natural selection, the question of the origins of life is to a large extent, quite independent of the question of the nature of its subsequent evolution. There are many different theories of the origins of life, consistent with the facts of evolution. These theories are highly speculative and hard evidence for or against them is difficult to find.[44]

I want to suggest that the situation is similar with regard to the origins of mathematics. It may well be that mathematics began as a perceptually warranted system of

[43] In particular, he has in mind the problem discussed by Benacerraf in 'What Numbers Could Not Be' [Benacerraf 1965].
[44] See for example [Cairns-Smith 1985].

beliefs, or perhaps more accurately, as *pragmatically* justified by its utility in solving practical problems of measurement and counting . What is far less clear is that the theory of the ideal-collector (or something like it) is strictly required to make good this claim.

It is not that I think that the theory of the ideal-collector must be wrong, it is only that I want to remain neutral, as far as possible, about such ontological questions. It does not seem to me that answering them is a prerequisite for answering the sorts of epistemological questions I am interested in. We need not worry too much about the origins (perceptual or not) of mathematics, just as evolutionary biologists do not worry too much about the origins of life. As we shall see, there are problems enough in giving an account of the nature of evidence in mathematics, problems which are independent of its origins.

Kitcher's account is aimed at showing how our mathematical *knowledge* is acquired. He asks 'how do we know the mathematics that we do?' and answers 'by acquiring it from authorities whose knowledge is derived by extending in rational ways the knowledge of previous authorities'. One part of Kitcher's task then is to provide an account of knowledge; an answer to the question when in general a belief should count as being known by someone. Kitcher's general schema for answering this question is that knowledge is *appropriately warranted true belief* [ibid. p. 17]. The remaining task is that of describing the nature of the appropriate warrants in mathematics. Kitcher argues that perception can warrant belief in some primitive mathematics and that the inter-practice transitions transfer warrant from one practice to its extension.

In chapter two, I argued that the epistemological problem in the philosophy of mathematics is independent of any theory of knowledge. For me, the problem is not 'how do we *know*?' but 'what is the *evidence*?' According to the kind of general account of

knowledge which Kitcher prefers, evidence or justification is not necessary for knowledge; one can know something, on such a view, without being able to explicitly formulate any argument which would justify the belief. A true belief may be warranted by a process which is reliable, even though the subject can give no account of the reliability of that process. They may even be unaware that the process is in fact reliable. In such cases, the subject would have knowledge without evidence. This may be why Kitcher does not make much explicit use of the concept of evidence in his epistemology for mathematics.

Nonetheless, justification and evidence are surely a very important kind of warrant for beliefs. They are especially important in mathematics and the other sciences. Marcus Giaquinto makes this point in his paper on visualization as a source of knowledge in analysis [Giaquinto 1994]. He endorses a reliabilist account of knowledge, but points out that this kind of account is 'quite consistent with the fact that demonstrable justification is indispensable in any collective endeavour to augment knowledge.' [ibid. p. 792]. This point applies very clearly to Kitcher's account. For Kitcher, much of our mathematical knowledge is acquired by the testimony of authorities. But this sort of process is clearly not one which can warrant belief in the absence of someone being able to cite evidence for that belief; evidence which can be expressed in a public language and assessed by others against appropriate standards. As Giaquinto remarks:

> If we are to rely on some claim announced by a Galois without rediscovering it ourselves, it is not enough that he has discovered it (and so know[s] it): we must know that it has been discovered, and this almost always requires that someone has seen and checked a justification for believing it. This is why proof is so important in mathematics.
>
> [ibid. p. 792]

Indeed, the inter-practice transitions which Kitcher discusses are generally not of a kind which warrant belief in a way that does not require evidence. In fact, although Kitcher refers to them as 'patterns of change', they are really *kinds of evidence*. For example, one kind of evidence we can have for a mathematical theory is that we can use it solve problems – this is justification by question-answering. Another kind of evidence is that theory provides an explanatory systematization or generalization of prior results. That a theory provides a rigorous reconstruction of previously unrigorously obtained results is also a kind of evidence for the theory. Here we find a common pattern; a modification to mathematical practice is justified in terms of its consequences. I shall look at this general pattern of justification in more detail in chapter five.

We should certainly leave open the possibility that there may be ways of acquiring mathematical knowledge which do not require demonstrable justification, but we ought also to acknowledge the central role that such justification plays in mathematics. Kitcher is quite right to point out that mathematicians are involved in a communal effort to increase mathematical knowledge. That being so, it is obvious that knowledge acquired by an essentially private, if reliable, belief forming mechanism is largely irrelevant – what counts is knowledge which can be communicated to others.[45]

In subsequent chapters I will be attempting to answer the question 'what kinds of justification or evidence do we have for our mathematical beliefs?' It is then a further,

[45] The great Indian mathematician Ramanujan, who came to Cambridge under the auspices of G. H. Hardy seems to have had such a mechanism. He claimed that some of his theorems came to him in dreams, in which he would see scrolls with mathematical formulae written on them. On waking, he would simply write them down. This belief forming process seems to have been astonishingly reliable – Ramanujan filled notebook after notebook with statements nearly all of which have since been proved to be correct. It may well be that Ramanujan knew without having what we would call evidence. Nonetheless, Hardy apparently took great pains to stress to him the importance of giving proofs. Part of the expl nation of Ramanujan's failure to understand the importance of proof might be his initial isolation from the mathematical community; his mathematics was initially something done by and for himself alone and so ok a while for him to recognise the necessity of providing *evidence* for his theorems. See [Kanigel 1991].

perhaps not especially important question, to ask if the justification we have for our mathematical beliefs are such as to provide us with knowledge.[46] Only the latter question requires an account of knowledge; the former question, the one that will concern me here, does not.

Although Kitcher takes a broadly descriptive approach to the epistemology of mathematics, he does want to show that the kinds of inter-practice transitions he has identified are *rational*. There are several ways in which one might go about this:

(1) Show that the patterns of inference to be found in mathematics are analogous to other patterns of inference, already assumed to be rational; inferences found in the natural sciences for example.

(2) Show how the patterns of inference can be explained in the light of an appropriate metaphysical account of the subject matter of mathematics; for example, by providing an ontological interpretation of mathematics which shows how there can be perceptual evidence for mathematics.

(3) Show how the patterns of inference instantiate patterns which are rational according to some previously established general theory of rationality (or 'scientific method'). For example, one might attempt to show how they conform to valid inferences in some inductive logic.

[46] Indeed if an account of knowledge suggested that our justifications were not adequate for our mathematical beliefs to count as knowledge, this might well be taken as showing that account itself to be inadequate, rather than showing that we do not, after all, have any mathematical knowledge.

Kitcher explicitly rejects method (3). In agreement with many contemporary philosophers, he doubts that any such general theory can be found in advance of the project of giving a descriptive account of rational inference. For how would such a general theory be supported? Only by being in agreement with our actual inferential practice:

> Epistemology has no Archimedean point from which it can exert leverage on the knowledge claims of those who participate in the various kinds of human inquiry. A full account of what knowledge is and of what types of inferences should be counted as correct is not to be settled in advance. Rather, it must emerge from consideration of the ways in which humans actually infer and from the knowledge claims we actually make.
>
> [ibid. p. 97]

For the most part, Kitcher makes extensive use of the first strategy. For example, the rationality of systematization in mathematics is defended by pointing out how the benefits which such systematization brings are exactly analogous to the benefits seen to accrue to systematic theories in the natural sciences; the ability to deduce a wide range of diverse phenomena from a few basic principles [ibid. p. 218-9]. Kitcher does make occasionally use of the second strategy; sometimes he brings in his theory of the ideal collector in order to cast light on the rationale behind some inter-practice transitions. It is worth noting however that this strategy is never used by itself. Kitcher always gives an argument of the first kind before to attempting to show how his preferred ontology can provide some additional insight.

This comes out quite clearly in Kitcher's discussion of reinterpretation in mathematics. He wants to explain why mathematical theories seem to have a far higher rate of survival than scientific theories. He advances the claim that mathematical theories are often saved from refutation by reinterpretation. A mathematical theory that would be

refuted by some item of evidence is saved by reinterpreting it in a such way that the evidence ceases to be relevant.[47] But why is this sort of reinterpretation rational? Why is not just an *ad hoc* move designed only to save a theory from refutation? Kitcher gives two arguments which are intended to show how this sort of change can be rational. The first argument makes use of method (1):

> ...the root idea is readily comprehensible in terms of a division of labor which began in ancient science. Initially, mathematics included optics, astronomy, and harmonics as well as arithmetic and geometry....What has occurred since is a continued process of dividing questions among specialists. The old mathematical investigations of light, sound and space are portioned into explorations of the possibilities of theory construction (the province of the mathematician) and determinations of the correct theory (the province of natural scientists). This division of labour accounts for the fact that mathematics often resolves threats of competition by reinterpretation, thus giving a greater impression of cumulative development than the natural sciences.
>
> [ibid. p. 159]

I am not concerned here with whether or not Kitcher's argument here is correct. I want only to point out the form of the argument – the practice is rational because it is an example of something – the division of intellectual labour – which we agree to be rational in other cases. Hence this is an application of strategy (1). In the following paragraph, Kitcher then applies strategy (2), he attempts to use his ontological account of mathematics to further defend the claim that the practice is rational:

> Consider this practice in light of the picture of mathematical reality advanced in the last chapter. Mathematics begins from studying physical phenomena, but its aim is to delineate the structural features of those phenomena. Our early attempts to produce mathematical theories generate theories which, we later discover, can be amended to yield theories of comparable richness and articulation.

[47] The most obvious example, which is discussed in some detail in the next chapter, is the case of Euclidean geometry. For Kitcher's discussion see [Kitcher 1984, pp. 158-61].

> When this occurs, we regard both the original theory and its recent rival as concerned with different structures, handing over to our scientific colleagues the problem of deciding which structure is instantiated in the phenomena we set out to investigate.
>
> [ibid. pp. 159-160]

This second argument does very little work over and above that already done by the first argument. In particular, notice that the details of the ideal collector theory play no role in the second explanation. A structuralist or platonist could make exactly the same remarks to much the same effect. The second argument also carries far less conviction than the first. I am far more convinced that the division of labour which Kitcher points to is rational than I am that *any* particular ontological account of mathematics is correct. This point is generally applicable; an argument of the first kind will always be preferable to one of the second kind, because we can generally be far more confident of the premises of an argument of type (1) than we can be of premises which depend on the correctness of an account of the ontology of mathematics. This is indicative of the fact that a large class of problems in the epistemology of mathematics are quite independent of the nature of its ontology. I suggest then, that just as we can do without method (3), we can also do without method (2).

Should we in fact, give up any attempt to argue that the methods we use are rational? One might suppose that since to give an argument is to make use of the methods of rational inference we are trying to justify, any such attempt is open to the threat of circularity. However, although it is true that we could never have a global justification for *all* of our methods of rational inference, this does not mean that we cannot provide local justifications for a particular inferences. We might for example, show how the particular

inference is of the same form as inferences found in other fields of inquiry and already assumed to be rational. That is, we still have the first method at our disposal.[48]

Furthermore, although method (3) as described above seems hopeless, this does not mean that no general theory of rational inference is possible, against which we can judge particular inferences. What may be possible is to give a general, descriptive account of patterns of rational inference which can be used to explain and criticise particular cases. The difference between this and method (3) is that we give up the idea that we can arrive at such a theory *in advance*, before attempting to describe the inference we actually make and which we take to be rational.[49]

Clearly there are two different problems here. The first problem is to give an adequate descriptive account of the patterns of inference and cannons of evidence to be found in mathematics. The second problem is one of showing that those patterns are rational; either by showing that they are analogous to similar patterns already assumed to be rational, or by fitting them into a general descriptive theory of rational inference. It is the first problem which is fundamental; obviously there can be no question of justifying evidential practice in mathematics without an adequate description of that practice. I will not be concerned with the second problem here. My aim is only to contribute something towards the first problem, by giving a descriptive account of the nature of evidence in mathematics.

[48] Of course this would only be a relative or conditional form of justification: if this form of inference is rational, then so is this particular inference. So method (1) cannot completely avoid the charge of circularity. But it may be, as Goodman argues that this kind of justification is the best we can ever hope to achieve See [Goodman 1983, p. 64].

[49] Maddy has recently argued that there might be something at least to method (3). That is, we might make use of a very minimal means-end theory of rationality; we can identify the goals of mathematical practice and show how an element of the practice is justified by the way it achieves those goals, then: '..the judgement that the arguments depicted in the model are rational is based on a simple fundamental of practical reason: the soundness of means-end reasoning.' [Maddy 1997, p. 197]. Maddy applies this strategy very effectively to the question of the justification of candidates for new set-theoretic axioms.

7. Descriptive Epistemology

I should like to conclude by drawing attention to three methodological principles which have emerged from the discussion so far. These are the principles that an epistemology for mathematics should be (i) descriptive rather than normative, (ii) sensitive to practice and history and (iii) ontological neutral.

Consider (i) first. This is implicit in the work of Lakatos and quite explicit in that of Kitcher. In their work, there is no attempt to provide mathematics with an extra-mathematical or extra-scientific justification; certainly there is no attempt to provide mathematics with epistemologically secure foundations.[50] Instead, what we find is a detailed attempt to describe the modes of justification at work in mathematics.

The second principle concerns the ways in which any such descriptive account of mathematics is itself to be justified. As I argued in chapter two, we want our account of the epistemology of mathematics to be consistent with and hopefully illuminate actual mathematical practice. We do not want an account which merely shows one way in mathematics *could* be justified, we want it to tell us how mathematics actually *is* justified. As Kitcher and Lakatos both realise, the way to test epistemological accounts of mathematics in this respect is to play close attention to the history of mathematics. That history provides us with a rich source of examples of the varieties of evidence in mathematics, examples we can use to develop and test our theories of the epistemology of mathematics.

[50] Kitcher's use of method (1) for example clearly does not involve any attempt to provide mathematics with an extra-scientific justification, for the justification assumes that the methods of the natural sciences are themselves rational.

The third principle is that of ontological neutrality. This is the idea that the project of giving a descriptive account of the epistemology of mathematics which illuminates the actual practice and historical development of the subject is independent of the metaphysical problem of giving an account of the subject matter of mathematics. Lakatos provides us with no theory of the ontology of mathematics and his account of its development does not seem to require one. Kitcher does provide such a theory, but as I have argued, its role in his account of the origin and growth of mathematical knowledge may be eliminable. It is certainly not one which could not be played by any one of a range of other accounts.

This is not to say that ontological questions are of no philosophical interest. Any adequate philosophy of mathematics should provide answers to both the ontological question – what mathematics is about - and the epistemological question - how we know the mathematics that we do. The descriptive approach holds out the hope that we can make significant progress with the epistemological problem without having to solve the ontological problem before we can begin.

These then are the guiding principles of what I have been calling *descriptive epistemology*; the project of giving an ontologically neutral, descriptive account of the epistemology of mathematics; an account which is sensitive to practice and history. The idea is not of course an entirely new one. Aspects of it are to be found in the writings of many philosophers and mathematicians. Apart from Kitcher and Lakatos, we can find similar themes in the work of Gödel [1947], Wittgenstein [1933,1953], Quine [1969], Hilary Putnam [1971], Nelson Goodman [1983], Stuart Shapiro [1989], Ian Stewart [1987] and Reuben Hersh [1979, 1981] to name but a few. An interesting collection of papers which apply this approach to the history and methodology of mathematics is to be found in

History and Philosophy of Modern Mathematics [Aspray and Kitcher (eds.) 1988] which also contains a useful historical introduction.

Perhaps the most explicit formulation of descriptive epistemology in its application to mathematics however, is to be found in Penelope Maddy's recent book *Naturalism in Mathematics* [Maddy 1997]. Maddy draws on remarks of Gödel and Wittgenstein and the writings of Quine on 'naturalized epistemology' to develop an approach to the epistemology of mathematics which she calls *naturalism* and which has strong affinities to my descriptive epistemology. Maddy's aim in the book is to give an account of the methodology of set-theory; in particular, she is concerned with the status of the independent questions such as the continuum hypothesis and the arguments for and against candidates for new set-theoretic axioms which could settle them[51]. Since she is primarily concerned with contemporary set theory, she does not emphasise the historical development of mathematics. But she does argue forcefully that the account of the epistemology of mathematics we are aiming for should be consistent with and illuminate mathematical practice, that it should be descriptive rather than normative and independent of metaphysics.[52]

In the remaining chapters of the present work I will be discussing some problems of descriptive epistemology. My aim is to apply this approach to the investigation of the

[51] I will discuss some of the features of the justification of set-theoretic axioms which Maddy uncovers from the perspective of her naturalism in chapter six.
[52] For example, Maddy writes: '[t]o judge mathematical methods from any vantage-point outside mathematics, say from the vantage-point of physics, seems to me to run counter to the fundamental spirit that underlies all naturalism: the conviction that a successful enterprise, be it science or mathematics, should be understood and evaluated in its own terms, that such an enterprise should not be subject to criticism from, and does not stand in need of support from, some external, supposedly higher point of view. What I propose here is a mathematical naturalism that extends the same respect to mathematical practice ... my naturalist takes mathematics to be independent of both first philosophy and natural science.....in short, from any external standard.' [Maddy 1997, p. 184]. See also [ibid. p. 133-160, 190-1] on the importance of sensitivity to mathematical practice and the irrelevance of ontology to the methodology of mathematics.

various kinds of evidence to be found in mathematical inquiry. I would like to begin by examining Quine's claim that there can be empirical evidence of a certain kind for mathematics and that indeed, this is the *only* kind of evidence we have for our mathematical beliefs.

CHAPTER FOUR

EMPIRICAL EVIDENCE

Can there be empirical evidence for mathematics, evidence which derives ultimately from the testimony of the senses? Many philosophers have thought there cannot be. There at two main features of mathematics which provide support for this view. The first is the seemingly highly abstract nature of the subject matter of mathematics, its concern with objects that are quite remote from the world of sense experience. The second feature concerns the methodology of mathematics, at least at first glance there appears to be a distinct lack of empirical methods in mathematics. Mathematicians give various arguments for their theorems, but they do not carry out experiments or base their mathematical conclusions on the results of observations.

In fact, mathematics has always posed a problem for empiricism. If all our knowledge is at root empirical and if there can be no empirical evidence for mathematics, then it seems as though we have no reason for thinking that mathematics is true at all. Few empiricists however, have been willing to simply bite the bullet and say that we should simply reject mathematics, at least insofar as it has any claim to be a body of truths.[1]

Mill, it will be recalled, solved this problem by denying both features of mathematics mentioned above. Arithmetic, in his view, is not concerned with abstract objects, but is a body of truths about the results of counting and operating on collections of physical objects. Statements like '2 + 2 = 4' are simply abbreviated statements of highly

[1] The exception of course, would be those empiricist who adopt a formalist account of mathematics.

confirmed inductive generalisations, gained from our experience of counting collections of physical objects.[2]

An alternative kind of solution has been a fairly constant thread in the history of empiricist thought. This is the idea that mathematics is a body of truths concerning the relationships between our concepts, or in more recent times, their linguistic representations. The view of the logical positivists is a good example. The positivists wanted to say that a statement was cognitively significant, or meaningful, only if it was capable of being empirically confirmed or refuted. In this way, they hoped to exclude the more extreme varieties of speculative metaphysics. But in the process, they also seemed to exclude mathematics and logic, to which empirical evidence seems irrelevant. The claim that mathematics is a body of *analytic* truths was their response to this problem; a statement is cognitively significant if its either empirically verifiable or analytic. An analytic statement is one which is true in virtue of the meanings of its component terms. '2 + 2 = 4' for example is true because of the stipulations we have laid down governing the use of the symbols '2', '4', '+' and '='. As such it is completely devoid of empirical content and this explains the irrelevance of empirical evidence to mathematics. No fact about the world can contradict the statement '2 + 2 = 4', because its truth does not depend on facts about the world, but only on facts about what the mathematical symbols occurring in it mean.[3]

In the 1950s, W.V Quine launched an attack on this 'dogma of empiricism', the idea that the distinction between analytic and synthetic statements can be made to do useful work in philosophy. In particular, he argued that the analytic-synthetic distinction cannot be

[2] See [Mill 1884].
[3] See for example [Ayer 1936].

made precise in a way that would allow it to play the role which empiricists had long found for it, that of accounting for our knowledge of mathematics and logic.[4]

But if the analytic-synthetic distinction cannot play this role, we are back to the old problem of explaining how our knowledge of mathematics can be consistent with empiricism. Again, the apparent abstractness of mathematics seems to cast doubt on the possibility of providing it with an empirical justification. As I argued in chapter two, this is the real force of Benacerraf's epistemological challenge to platonism. It is not the causal theory of knowledge but *empiricism* which seems to make mathematical knowledge impossible. Since abstract objects are not physical and not causal agents, they cannot be perceived in any way. So how can there be empirical evidence for them, if empirical evidence is grounded ultimately in perception?

I will not be concerned here with the details of Quine's arguments against the analytic-synthetic distinction. What will concern me is the new solution to the problem of mathematics which Quine adopted. Quine's genius was to see that the problem could be solved by denying something that nearly everyone before him had taken for granted, namely the thesis that empirical evidence is irrelevant to mathematics, without having to give up the thesis that mathematics is concerned with abstract objects.

1. The Web of Belief

Towards the end of *Two Dogmas of Empiricism* Quine describes the metaphor of the web of belief:

[4] [Quine 1951]. See also [Quine 1936, 1948, 1957, 1962].

> The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or to change the figure, total science is like a field of force whose boundary conditions are experience.
>
> [Quine 1951, p. 42]

Quine argues that the mistake made by earlier empiricists was to think that individual statements can be evaluated empirically in isolation from each other. Instead, we have to recognise that our scientific beliefs form an interlocked system or web, which "faces the tribunal of sense experience ... as a corporate body." [ibid. p. 41]. For Quine, the ultimate justification of the system lies in its ability to help us predict and explain the flow of sensory experience.

Quine recognised that if this picture is correct, then we can account for our knowledge of mathematics. For mathematical beliefs form an indispensable part of the system of total science. Hence, they are justified to the extent that they contribute to the attainment of the goals of scientific prediction and explanation. In Quine's view, the mathematics used in a successfully confirmed scientific theory is confirmed along with the rest of that theory. Mathematical statements should be thought of as empirical hypotheses like any other – hypotheses which are indispensable in providing us with simple and powerful theories which allow us to predict and organise the data of sense experience. On this view, mathematical objects like numbers and sets are theoretical posits, on a par epistemologically speaking with electrons and photons:

> As an empiricist, I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience...physical objects...are posits which serve merely to simplify our treatment of experience... Objects at the atomic level and beyond are posited to make the laws of macroscopic objects, and ultimately the laws of experience simpler and more

manageable...Physical objects, small and large, are not the only posits....the abstract entities which are the substance of mathematics – ultimately classes and classes of classes and so on up – are another posit in the same spirit.

[ibid. pp. 45-6]

Quine draws another conclusion. If mathematics can be supported by empirical considerations, it can also be refuted by them:

Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune from revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?

[ibid. p. 43]

If there is evidence which disconfirms a scientific theory, such evidence can also disconfirm the mathematics used in that theory. Our mathematical beliefs are not certain or necessary, they are open to empirical confirmation, falsification and revision in the same way as our scientific beliefs. The illusion of a difference between mathematical and other statements in this regard is generated, according to Quine, by purely pragmatic considerations. We are less inclined to revise the mathematical and logical components of our theories because they are so deeply embedded in the system of total science that altering them would result in a massive restructuring of that system. But where revision of mathematical or logical statements would result in an overall simplification or improvement in the total system of science, it is an option that is open to us.

Quine's picture is attractive. It provides us with an account of mathematical knowledge which is consistent with empiricism and it does so without any reinterpretation of mathematical discourse. In this way, Quine's epistemology for mathematics comes the closest to providing us with a solution to Benacerraf's dilemma, as I formulated it in chapter two. In particular, the epistemological challenge to platonism is met. Mathematical objects are indeed abstract, but their utility in enabling us to predict and explain the world provides us with all the justification for their existence we need or could ever have.

Quine's picture is not without critics however. His account of the epistemology of mathematics has been criticised on the grounds that it does not accord very well with mathematical practice.[5] In particular, Quine's account appears to overestimate the extent to which mathematical theories can be confirmed or disconfirmed by empirical evidence. Mathematicians do not in practice take the indispensability of mathematics in scientific theories as the sole standard by which mathematical theories are to be evaluated. Quine's epistemology seems to leave a great deal of mathematics unaccounted for. Those portions of mathematics (the higher flights of set theory for example) which are not applied in scientific theories receive no justification at all by Quine's argument. Quine's attitude is to bite the bullet and say that unapplied mathematics should not be thought of as on an equal footing with empirically confirmed mathematics. There is in fact, no reason to think that any of it is true. This is not to say research in such areas of mathematics is pointless; after all what is unapplied now may well turn out to be just what the physics of the 22nd century needs. But we do not need to take the ontological commitments of such mathematics seriously; mathematics of this kind is best viewed as an investigation into what would follow *if* certain axioms are true, but in the absence of any application of the theory to the

[5] See for example [Maddy 1992, 1997].

serious business of science, there is no reason for thinking that the axioms *are* true.[6] But again, this conclusion seems to conflict with the practice of mathematics, where we find arguments for and against mathematical principles (powerful new axiom candidates for set-theory for example) which do not appeal to the use of such principles in scientific theories.[7]

However, Quine does not simply describe an attractive metaphor. He also has an argument for his epistemology. This is the famous *indispensability argument*. That argument concludes that there is empirical evidence of a certain kind for mathematics. This conclusion is quite consistent with there being *other*, non-empirical kinds of evidence in mathematics. Quine does make a further claim however, that empirical evidence of this kind is the *only* evidence we have for mathematics. Hence his dismissal of mathematics which plays no role in empirically well confirmed theories. This second claim depends not on the indispensability argument but on Quine's empiricism or *naturalism*; ordinary empirical evidence of the kind which justifies mathematics and science is the only kind of evidence we can legitimately ask for; there is no higher standard by which our beliefs can be judged.[8]

The arguments against Quine's epistemology for mathematics which appeal to mathematical practice, if they succeed at all, count against this second claim, rather than the first, since the features of practice alluded to suggest that there are other, apparently non-empirical kinds of evidence in mathematics. What I want to do here is consider first the question whether Quine's indispensability argument succeeds in establishing his first claim, that there is empirical evidence for mathematics. To anticipate, I will argue that the

[6] See [Quine 1984, p. 788, Putnam 1971, p. 56].
[7] See [Maddy 1997, pp. 158-60].
[8] See [Quine 1969a, 1975, Putnam 1971, pp. 71-4].

indispensability argument faces certain difficulties, connected with the problem of giving an account of the relationship of evidence to theory. Nonetheless, there is a way of stating the argument which, though not immune to these problems, does stand some chance of success. Hence, I conclude with a qualified endorsement of Quine's first claim, that there can be empirical evidence for mathematics. However, it is a further question what role this kind of evidence plays in mathematical methodology, that is, whether empirical evidence is the only kind of evidence we have for mathematics. I shall return to this issue in the concluding section of this chapter. Let us begin however, by taking a closer look at Quine's argument.

2. QUINE'S ARGUMENT

Statements of pure mathematics are indispensable in deriving the empirical predictions which confirm or disconfirm scientific theories. They are therefore confirmed or disconfirmed by those predictions. Mathematics is thus open to confirmation and revision in the light of empirical evidence. This is the analysis of Quine's indispensability argument given by W.D Hart in his paper 'Access and Inference' [Hart 1996]. According to Hart, the argument is based on two premises:

> I call the first premiss Duhem's thesis: beyond a minimal level, no scientific hypothesis is tested individually; only relatively large and heterogeneous bodies of hypotheses are tested against experiment and observation....I call the second premiss of Quine's argument his 'indispensability thesis': beyond a minimal level, we do not know how to do natural science without mathematics...any reasonably sophisticated natural scientific theory could be formalized only as an extension of some part of mathematics.....It follows from Duhem's thesis and Quine's indispensability thesis that such mathematics as is required by natural science for making true predictions (and by Quine's thesis, such mathematics there is) is confirmed thereby; there are no grounds for confining such confirmation to the

natural science and denying it to the mathematical science. I call this conclusion of Quine's argument his epistemology for mathematics.

[Hart 1996, pp. 52-53]

Duhem's thesis, or *holism* in Quine's terminology, has become a commonplace in the philosophy of science. But why should we think that 'only large and heterogeneous bodies of hypotheses are tested against experiment and observation'? A common answer (and the answer Hart provides) is to appeal to the idea that theories are confirmed and disconfirmed by deriving testable predictions from them. However, as we saw in our discussion of Popper's philosophy of science, theories do not entail observations by themselves; it usually requires a large number of auxiliary hypotheses to derive a testable prediction from a theory. It follows that only such large numbers of hypotheses are confirmed or disconfirmed by the predictions they entail. Here is Quine's version of this argument for holism in *Pursuit of Truth*:

> The observational test of scientific hypotheses....consists in testing observation categoricals that they imply...In order to deduce an observation categorical from a given hypothesis, we may have to enlist the aid of other theoretical sentences and of many common-sense platitudes that go without saying, and perhaps the aid even of arithmetic and other parts of mathematics. In that situation, the falsity of the observation categorical does not conclusively refute the hypothesis. What it refutes is the conjunction of sentences that was needed to imply the observation categorical. In order to retract that conjunction we do not have to retract the hypothesis in question; we could retract some other sentence of the conjunction instead. This is the important insight called Holism.
>
> [Quine 1992 pp. 12-13]

Holism then, is based on the idea that scientific theories and hypotheses are confirmed or disconfirmed by deriving testable consequences from them. This idea is known as the *hypothetico-deductive* account of confirmation. Holism follows from the H-D account and a simple minor premise. The H-D account states that if a conjunction of hypotheses entails a correct prediction, then the conjunction is confirmed by that prediction. If a conjunction of hypotheses entail a false prediction, then the conjunction is disconfirmed by that prediction. Since it typically requires a large number of hypotheses to derive a testable prediction, it follows that only large numbers of hypotheses are confirmed or disconfirmed by the predictions they entail.[9]

Consider now the indispensability thesis. This is the claim that mathematics is indispensable in empirical science as it is now practised. This means that mathematical statements will typically be included in the large body of hypotheses required to derive a testable prediction from a scientific theory.

The H-D account of confirmation and the indispensability thesis entail that a true prediction confirms the conjunction of hypotheses used to derive that prediction and that these hypotheses will include some mathematical statements. If the prediction is false, it will disconfirm a conjunction of hypotheses which includes mathematical statements.

However, the conjunction of hypotheses used to derive the prediction is not itself a mathematical statement, so we cannot yet conclude that there can be empirical evidence for any purely mathematical statement. We can only conclude that such a prediction confirms a conjunction of statements which includes some mathematics. In order for Quine's argument to go through, we need an additional premise, which I call the *distribution principle*: If a prediction confirms a conjunction of hypotheses, then the prediction confirms each conjunct of that conjunction.

[9] The use of the term *prediction* in this context is not meant to imply that the consequences of a theory which confirm or disconfirm it must be statements of the occurrence of some future event. For Quine, they must ultimately be observation categoricals - statements capable of more or less direct observational confirmation. In what follows however, I will not make any special assumptions about the predictions taken to confirm theories.

The hypothetico-deductive account of confirmation and the distribution principle together entail that empirical evidence can confirm mathematical statements. Let $\alpha\mathbf{C}\beta$ stand for 'α confirms β' and let $\alpha\mathbf{D}\beta$ stand for 'α disconfirms β'. The H-D account of confirmation can then be expressed as follows:

(HD) $\alpha_1 \,\&\, \alpha_2 \ldots \&\, \alpha_n \vdash \beta \;\Rightarrow\; \beta\mathbf{C}(\alpha_1 \,\&\, \alpha_2 \ldots \&\, \alpha_n) \,\&\, {\sim}\beta\mathbf{D}(\alpha_1 \,\&\, \alpha_2 \ldots \&\, \alpha_n)$

In words, (HD) says that if a conjunction of sentences entail a prediction β, then the prediction confirms the conjunction and the negation of the prediction disconfirms the conjunction. Strictly speaking, the schema (HD) defines a relation of *possible*, rather than *actual* confirmation. If a theory entails a prediction β, then β is possible confirming evidence for that theory. If β is in fact true, then the theory is actually confirmed by β. Likewise for disconfirmation, if a theory entails a prediction β, then ${\sim}\beta$ is possible disconfirming evidence for the theory. If β is in fact false, then the theory is actually disconfirmed by the prediction.

Notice that (HD) states only a sufficient condition for confirmation. No claim is being made that entailing a prediction is necessary for confirmation. It may well be for example, that a theory can be confirmed by evidence which it does not entail but on which it confers a high degree of probability. All that is required for Quine's indispensability argument however, is that entailing a prediction is a sufficient condition for confirmation.

The relation defined by (HD) is a relation of qualitative, rather than quantitative confirmation.[10] If you prefer to think in terms of degree of confirmation, then $\alpha\mathbf{C}\beta$ can be read as 'α confirms β to some degree', the exact amount of support which α confers on β being unspecified.

The distribution principle can be stated as follows:

(DIST) $\beta\mathbf{C}(\alpha_1 \,\&\, \alpha_2 \ldots \&\, \alpha_n) \Rightarrow \beta\mathbf{C}\alpha_1 \,\&\, \beta\mathbf{C}\alpha_2 \ldots \&\, \beta\mathbf{C}\alpha_n$

Again, the confirmation referred to here is purely qualitative. No claim is being made that evidence confirms each conjunct of a conjunction to exactly the same degree as it confirms the conjunction as a whole. In quantitative terms, the principle should be read as stating that if β confirms a conjunction to some degree, then it confirms each conjunct to some degree also, but not necessarily to the same degree.

(HD) and (DIST) together entail that any mathematical statement used in deriving a prediction from a theory or hypothesis is confirmed by that prediction. To see how this works, let us consider an example. Kepler's third law states that the orbital period of a planet is directly proportional to the square root of the cube of the planet's distance from the sun, measured along the major axis of the ellipse which describes the planet's orbit. We have:

(1) $$\forall p\left(\mathrm{T}(p) = k \times \mathrm{D}(p)^{3/2}\right)$$

where $\mathrm{T}(p)$ is the orbital period and $\mathrm{D}(p)$ the distance of a planet p from the sun. Since it takes the Earth one year to orbit the sun and the distance of the Earth from the sun is (by

[10] See [Hempel 1945a pp. 5-7].

definition) one astronomical unit, substitution into equation (1) gives a value for k of 1. So if we measure orbital periods in years and distance in astronomical units, equation (1) becomes simply:

(2) $$\forall p\left(\mathrm{T}(p) = \mathrm{D}(p)^{3/2}\right)$$

Equation (2) can be used to derive a testable prediction about the orbital period of any planet, given the planet's distance from the sun. I will take Mars as an example and set out the required deduction.

α_1:	$\forall p(\mathrm{T}(p) = \mathrm{D}(p)^{3/2})$	Kepler's third law.
α_2:	$\mathrm{T(MARS)} = \mathrm{D(MARS)}^{3/2}$	From α_1
α_3:	$\mathrm{D(MARS)} = 1.52$	Auxiliary hypothesis.
α_4:	$\mathrm{T(MARS)} = 1.52^{3/2}$	From α_2 and α_3
α_5:	$1.52^{3/2} = 1.87...$	Mathematical statement.
β:	$\mathrm{T(MARS)} = 1.87...$	From α_4 and α_5

Thus Kepler's third law, in conjunction with an auxiliary hypothesis and a mathematical statement, entails that Mars orbits the sun approximately every 1.87 years, or 683 days. This value is in good agreement with observation, so Kepler's third law is confirmed by it. We have, by the above derivation shown that:

(1) $$\alpha_1 \,\&\, \alpha_2 \,\&\, \alpha_3 \,\&\, \alpha_4 \,\&\, \alpha_5 \vdash \beta$$

By (HD) then, we can infer:

(2) $$\beta\mathbf{C}(\alpha_1 \,\&\, \alpha_2 \,\&\, \alpha_3 \,\&\, \alpha_4 \,\&\, \alpha_5)$$

By the distribution principle, we have:

(3) $$\beta\mathbf{C}\alpha_1 \,\&\, \beta\mathbf{C}\alpha_2 \,\&\, \beta\mathbf{C}\alpha_3 \,\&\, \beta\mathbf{C}\alpha_4 \,\&\, \beta\mathbf{C}\alpha_5$$

So β confirms not only α_1, Kepler's third law, but also the mathematical statement α_5, that $1.52^{3/2} = 1.87...$ Any more general mathematical statement which entails α_5 would also be confirmed by β according to the H-D account. For example, suppose μ_1 & μ_2 ...& μ_n entail α_5. The μ_i might be axioms of an appropriate mathematical theory, the axioms of Z-F set theory for example. Since α_1 & ... α_4 & α_5 entail β, it follows that μ_1 & ... μ_n & α_1 & ... α_4 entail β. Then, by (HD), β confirms μ_1 & ... μ_n & α_1 & ... α_4 and by the distribution principle, β confirms μ_1, μ_2 ... μ_n. In this way, Quine's construction allows for an empirical justification not only of the often quite specific mathematical statements (numerical equations such as α_5 for example) which are used to derive predictions from a theory, but also of any more general mathematical statements, such as set-theoretic axioms, which imply them.

3. Problems

Quine's epistemology for mathematics then, follows from three premises, the H-D account of confirmation (which entails holism), the indispensability thesis and the distribution principle. If we accept these premises, we must accept that empirical evidence is relevant to mathematics. If this conclusion is false, one or more of the premises must be given up.

In fact, as is fairly well known, the H-D account of confirmation is inadequate. It is wildly over-permissive, yielding a number of absurd results. We formulated the H-D account of confirmation and the distribution principle as follows:

(HD) $\alpha_1 \& \alpha_2 \ldots \& \alpha_n \vdash \beta \Rightarrow \beta\mathbf{C}(\alpha_1 \& \alpha_2 \ldots \& \alpha_n)$

(DIST) $\beta\mathbf{C}(\alpha_1 \& \alpha_2 \ldots \& \alpha_n) \Rightarrow \beta\mathbf{C}\alpha_1 \& \beta\mathbf{C}\alpha_2 \ldots \& \beta\mathbf{C}\alpha_n$

(HD) and (DIST) quickly lead to absurdity. For example (HD) entails that if a conjunction of hypotheses entail a prediction β, then that prediction confirms the conjunction of the original hypotheses with every statement γ. This is the so called *problem of irrelevant conjunction*.[11] We can express this problem using our notation as follows:

(IC) $\alpha_1 \& \alpha_2 \ldots \& \alpha_n \vdash \beta \Rightarrow \beta\mathbf{C}(\alpha_1 \& \alpha_2 \ldots \& \alpha_n \& \gamma)$ - for all γ

The derivation of (IC) from (HD) is quite simple:

(1)	$\alpha_1 \& \alpha_2 \ldots \& \alpha_n \vdash \beta$	Assumption.
(2)	$\alpha_1 \& \alpha_2 \ldots \& \alpha_n \& \gamma \vdash \beta$	From (1)
(3)	$\beta\mathbf{C}(\alpha_1 \& \alpha_2 \ldots \& \alpha_n \& \gamma)$	From (2) by (HD)

At line (2) we made use of the principle of *monotonicity;* logical entailment is preserved under addition of premises.[12] Notice also that since any logical truth is entailed by every statement, (HD) implies that any logical truth confirms every statement:

(LT) $\vdash\beta \Rightarrow \beta\mathbf{C}\alpha$

<u>Proof</u>

(1) $\vdash\beta$	Assumption.
(2) $\alpha \vdash \beta$	From (1)
(3) $\beta\mathbf{C}\alpha$	From (2) by (HD)

[11] See [Glymour 1980a, pp. 31-35].

[12] This is a fairly trivial consequence of the usual definition of logical consequence. A set of sentences Φ entails a conclusion α ($\Phi\vdash\alpha$) iff every interpretation of the language which makes every member of Φ true, also makes α true. Suppose then that $\Phi\vdash\alpha$. Let Ψ be any set of sentences. Then $\Phi\cup\Psi\vdash\alpha$. Proof: Let v be any interpretation which makes every member of $\Phi\cup\Psi$ true. Then v makes every member of Φ true. But since, by assumption, $\Phi\vdash\alpha$, any interpretation which makes every member of Φ true also makes α true. So in particular, v makes α true.

If we combine (HD) with the distribution principle, we can deduce that any arbitrary statement confirms every hypothesis:

(U) $\beta\mathbf{C}\alpha$ - for all β and α.

<u>Proof 1</u>

(1) $(\alpha \& \beta) \vdash \beta$	Logical Truth.
(2) $\beta\mathbf{C}(\alpha \& \beta)$	From (1) by (HD)
(3) $\beta\mathbf{C}\alpha \& \beta\mathbf{C}\beta$	From (2) by (DIST)
(4) $\beta\mathbf{C}\alpha$	From (3)

We can argue to the same conclusion as follows:

<u>Proof 2</u>

(1) $\alpha \& {\sim}\alpha \vdash \beta$	Logical Truth.
(2) $\beta\mathbf{C}(\alpha \& {\sim}\alpha)$	From (1) by (HD)
(3) $\beta\mathbf{C}\alpha \& \beta\mathbf{C}{\sim}\alpha$	From (2) by (DIST)
(4) $\beta\mathbf{C}\alpha$	From (3)

So (HD) and (DIST) entail the absurd conclusion that every observation confirms every hypothesis. The problem is one of relevance. Evidence confirms or disconfirms a hypothesis only if that evidence is relevant to the hypothesis. Logical entailment however, is notoriously insensitive to relevance.

For example, take our derivation of the orbital period of Mars from Kepler's third law. The prediction was entailed by a conjunction of sentences, $\alpha_1 \& \alpha_2 \& \alpha_3 \& \alpha_4 \& \alpha_5$, which included Kepler's third law, an auxiliary hypothesis and a mathematical statement. But, by the principle of monotonicity, we can add any arbitrary premise to the derivation and the deduction will still be valid. For example, let γ be the statement 'Isaac Newton died in the year 1666'. Then $\alpha_1 \& \alpha_2 \& \alpha_3 \& \alpha_4 \& \alpha_5 \& \gamma$ also entails that the orbital period of Mars is 1.87 years. According to (HD) then, that prediction confirms the conjunction and by distribution it confirms γ. But clearly data concerning the orbital period of Mars is simply irrelevant to the date of Newton's death.

It is not too difficult to see how (HD) can be amended to avoid these difficulties. Notice that in the proof of (IC) above, the irrelevant premise γ that we add in line (2) is *dispensable* in the proof of β from the conjunction α_1 & α_2 .... & α_n. Likewise, in proof 1, γ is dispensable in the proof of β from γ & β. In both cases, there is some proper subset of the conjunction which suffices to entail β. This suggests the following modified version of (HD):

(HD')
$$(\alpha_1 \,\&\, \alpha_2 \ldots \&\, \alpha_n \vdash \beta) \,\&\, (\alpha_1 \,\&\, \alpha_2 \ldots \&\, \alpha_n \nvdash) \,\&$$
$$\sim\exists\Psi(\Psi \subset \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \,\&\, \Psi \vdash \beta) \Rightarrow$$
$$\beta\mathbf{C}(\alpha_1 \,\&\, \alpha_2 \ldots \&\, \alpha_n)$$

In words, (HD') says that if a conjunction of hypotheses entail a prediction β and that conjunction is consistent and there is no proper subset of those hypotheses which entail β, then β confirms the conjunction.

(HD') neatly avoids all of the problems noted so far. The proof of (IC) is blocked at line (2). We cannot, using (HD') deduce from the fact that α_1 & α_2 .... & α_n & γ entails β that β confirms α_1 & α_2 .... & α_n & γ, because there is a proper subset of $\{\alpha_1, \alpha_2, \ldots, \alpha_n, \gamma\}$ namely $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ which (by assumption) suffices to entail β.

Likewise, the proof of (LT) is blocked at line (2). We cannot infer from the fact that α entails β, that β confirms α, because there is a proper subset of $\{\alpha\}$, namely the empty set $\{\}$, which suffices to entail the logical truth β.

In the first proof of (U), the proof is blocked at line (1). We cannot deduce from the fact that γ & β entails β that β confirms γ & β because there is a proper subset of $\{\gamma, \beta\}$, namely $\{\beta\}$ which entails β. Finally, in the second proof of (U), the proof is blocked at line (1), since the conjunction that entails the prediction β is inconsistent.

As originally stated, the indispensability thesis claimed merely that some mathematical statements will be *included* in the conjunction of hypotheses which entail a prediction. But we have seen that this is not sufficient for the prediction to confirm those hypotheses. We require not just that some mathematical statements be included in the hypotheses which entail the prediction, but that the mathematical statements be *indispensable* to the derivation of the prediction, in the sense that the prediction would not follow without them.

Notice that in the example given in section two (HD') is applicable, since every premise is indispensable to the derivation of the conclusion. So we would still get the result that the statement of pure mathematics, α_5, is confirmed by an item of empirical evidence. We also still have the result that any set of axioms, all of which are indispensable in entailing α_5, will also be confirmed by the same item of empirical evidence.

Notice how (HD') avoids the problem of irrelevant premises. One way in which a premise may be irrelevant to establishing a conclusion is when the premise is dispensable in deducing the conclusion. If irrelevant premises are always dispensable in this way, then (HD') will guarantee that evidence is relevant to the hypotheses which it confirms or disconfirms.[13]

[13] (HD') achieves this by ruling out certain deductions. Notice that the sorts of deductions it rules out are the sorts of deductions ruled out in relevance logic; deductions of an arbitrary statement from a contradiction and deductions in which one or more assumptions are not used. See for example [Read 1988, 1994, Haack 1978

However, indispensability in this sense is not sufficient for evidential relevance. Further difficulties quickly arise for (HD'). For example, from (HD') and (DIST) we can deduce the so called *Paradox of the Ravens* [see Hempel 1945a, pp. 13-15, Goodman 1983 p. 70]. The statement that a given thing is neither black nor a raven confirms the hypothesis that all ravens are black:

(RP) $(\sim Ba \,\&\, \sim Ra)\ \mathbf{C}\ \forall x(Rx \rightarrow Bx)$

<u>Proof</u>

(1) $\forall x(Rx \rightarrow Bx) \,\&\, \sim Ba \vdash (\sim Ba \,\&\, \sim Ra)$	Logical truth.
(2) $(\sim Ba \,\&\, \sim Ra)\ \mathbf{C}\ (\forall x(Rx \rightarrow Bx) \,\&\, \sim Ba)$	From (1) by (HD')
(3) $(\sim Ba \,\&\, \sim Ra)\ \mathbf{C}\ \forall x(Rx \rightarrow Bx)$	From (2) by (DIST)

The entailment in line (1) meets the conditions for (HD') to apply. The conjunction is consistent and there is no proper subset of the conjuncts which entails the conclusion. But the result is that an apparently irrelevant observation, that a certain object is neither black nor a raven, confirms the hypothesis that all ravens are black. This is indicative of a deeper problem. (HD') and (DIST) entail that an arbitrary observation confirms every sentence consistent with it, relevant or not:

(U') $(\alpha \,\&\, \beta) \nvdash\ \Rightarrow \beta\mathbf{C}\alpha$ - for all α and β

<u>Proof</u>

(1) $(\alpha \,\&\, \beta) \nvdash$	Assumption.
(2) $(\alpha \rightarrow \beta) \,\&\, \alpha \vdash \beta$	Logical Truth.
(3) $\beta\mathbf{C}((\alpha \rightarrow \beta) \,\&\, \alpha)$	From (1) and (2) by (HD')
(4) $\beta\mathbf{C}\alpha$	From (3) by (DIST)

pp. 197-203]. It might then be possible to reformulate (HD') so as to require that a conjunction of hypotheses must *relevantly entail* a prediction if that prediction is to confirm or disconfirm those hypotheses.

The entailment in line (2) again meets the conditions for (HD') to apply. The conjunction is consistent (by assumption) and neither conjunct is sufficient by itself to entail the conclusion. But once more we have an absurd result. Not every observation confirms every hypothesis consistent with it, since some consistent observations are irrelevant to a given hypothesis. The problem with (HD') is that any statement can be indispensable in entailing any item of evidence, given appropriate auxiliary hypotheses.[14]

I have been analyzing the hypothetico-deductive account in terms of confirmation rather than disconfirmation, but analogous difficulties arise for disconfirmation, since confirmation and disconfirmation are inter-definable. A sentence α disconfirms a sentence β if and only if α confirms the negation of β:

(Def **D**) $\alpha\mathbf{D}\beta \equiv \alpha\mathbf{C}{\sim}\beta$

Recall that (HD) and (DIST) entail that $\alpha\mathbf{C}\beta$ for all α and β. Thus we also have $\alpha\mathbf{C}{\sim}\gamma$ for all α and γ and therefore, by the definition of **D**, $\alpha\mathbf{D}\gamma$. Thus (HD) and (DIST) entail that every observation both confirms and disconfirms every hypothesis.

Likewise, since (HD') and (DIST) entail that if α and β are consistent, then $\alpha\mathbf{C}\beta$, it follows that if α and $\sim\gamma$ are consistent, then $\alpha\mathbf{C}{\sim}\gamma$. By the definition of **D**, $\alpha\mathbf{D}\gamma$. So (HD') and (DIST) entail that every observation confirms every hypothesis consistent with it and disconfirms every hypothesis whose negation is consistent with that observation.

It may be possible to formulate a better version of the H-D account of confirmation, one which avoids all these problems. However, the difficulties involved in providing an

[14] A.J Ayer in *Language, Truth and Logic* gives an exactly analogous criticism of this version of the H-D account of confirmation [Ayer 1946, pp. 15-18]. For a criticism of Ayer's amended account see [Church 1949].

account of confirmation and disconfirmation along these lines are acute.[15] Indeed, the work of Nelson Goodman seems to suggest that no purely syntactic account of confirmation, no account sensitive only to the logical form of the statements involved, is possible at all.[16]

4. Holism and Evidential Relevance

Evidence confirms or disconfirms a hypothesis only when that evidence is relevant to the hypothesis. The problem of characterizing the relationship between a hypothesis and its confirming or disconfirming evidence is then the problem of characterizing a certain relation of relevance, evidential relevance, between statements.

Is the empirical evidence which supports a scientific theory also relevant to the mathematics used by that theory? The indispensability argument we have been considering concludes that it is. The argument works by providing a certain partial definition of evidential relevance[17]. Evidence is taken to be relevant to a hypothesis if the hypothesis is an indispensable premise in entailing the evidence. We have seen, however, that this sort of indispensability is not sufficient for evidential relevance. A hypothesis may be indispensable in entailing evidence which is not relevant to it. The hypothetico-deductive account of evidential relevance on which Quine's argument is based is inadequate.

It might be objected that Quine's argument does not depend on the hypothetico-deductive account of confirmation or indeed on any account of evidential relevance at all. After all, we do have theories for which there is confirming evidence, even though the project of defining the relation which holds between the evidence and those theories is

[15] See for example [Hempel 1945a, pp. 97-101 and 1965a pp. 101-107, Glymour 1980a, pp. 29-48, 1980b].
[16] See [Goodman 1983 pp. 81-83; Putnam 1994b, pp. 394-399].
[17] Partial in that provides a sufficient but not a necessary condition for evidential relevance.

beset by difficulties. This suggests a version of the indispensability argument which does not depend on any particular account of evidential relevance. The argument would be that have good empirical evidence for our best confirmed scientific theories, mathematics is an indispensable part of those theories and therefore we have good empirical evidence for mathematics.

Let us consider an example of the application of this version of the indispensability argument. There is a body of evidence, consisting of the observed advance in the perihelion of Mercury, Eddington's measurements of the gravitational deflection of starlight and so on, which confirms the general theory of relativity. Applying the distribution principle, it follows that this body of evidence confirms every sentence of the general theory of relativity.[18] But the general theory of relativity includes many mathematical statements, so they too must be confirmed by that evidence.

Since it is plausible to suggest that all our best scientific theories can be formulated as extensions of some mathematical theory, exactly analogous arguments will show that any evidence for those theories is also evidence for the mathematics included in them. All we need for this argument to work is the distribution principle and an example of a mathematically formulated theory which is confirmed by some evidence. We do not need any particular account of the relation of theory to evidence.

This argument assumes that the empirical evidence for the general theory of relativity is evidence for the *whole* of that theory. In general however, this assumption is

[18] To apply the distribution principle (as formulated by (DIST) above) in this context, we would need to assume that the general theory of relativity can be thought of as a (possibly infinite) conjunction of sentences. The argument we are considering does not depend on this assumption however. Instead of the distribution principle, we can appeal to the principle that if α confirms a set of sentences Φ, then α confirms each member of Φ. Then if Φ is a theory (a deductively closed set of sentences) and α represents the empirical evidence for that theory, we can infer that α confirms every sentence Φ, including any mathematical statements which are elements of the theory.

false. A given item of evidence is seldom relevant to every part of a theory. Clark Glymour makes this point in *Theory and Evidence*:

> Recall the case of Kepler's laws...It seems that observations of a single planet (and, of course, the sun) might provide evidence for or against Kepler's first law (all planets move on ellipses) and for or against Kepler's second law (all planets move according to the area rule), but no observation of a single planet would constitute evidence for or against Kepler's third law (for any two planets, the ratio of their period equals the 3/2 power of the ratio of their distances).
>
> [Glymour 1980a, p. 84]

An item of evidence concerning a single planet is relevant to less than the whole of Kepler's theory, since such evidence, although relevant to the first and second laws, is not relevant to the third law.[19]

Scientific theories can be made up of a number of hypotheses which are independent of each other. In such cases, evidence which accrues to one hypothesis need not be relevant to any of the others. Kepler's theory of planetary motion is just one example. Another example comes from evolutionary biology – the 'neo-Darwinian synthesis'.[20] This theory can be thought of consisting of two main hypotheses. The first hypothesis is that evolution takes place by means of a certain kind of stochastic process, the process of natural selection identified by Darwin. The second hypothesis is that in all living organisms, the genetic information which this process requires is carried by the DNA molecule. These two hypotheses are independent. Suppose scientists discovered tomorrow that DNA is not after all the carrier of genetic information (perhaps they discover that some other molecule is responsible). Would we have to abandon evolutionary theory? Not

[19] In my example of the derivation of the orbital period of Mars from Kepler's third law in section two, it seemed as though the evidence did concern only one planet, namely Mars. But the auxiliary assumptions we used stated the distance of Mars from the sun in astronomical units and to calculate this we need data concerning the distance from the sun of another planet, namely the Earth.

[20] See [Dawkins 1986]

entirely. Obviously such evidence would count against the second hypothesis. But it would not disconfirm the first hypothesis. It would not show that species do not evolve through a process of natural selection, it would show only that we had been mistaken about the underlying molecular mechanism which implements the process.

If, in general, evidence can be relevant to less than the whole of a scientific theory, then the version of the indispensability argument we have been considering does not go through. If the empirical evidence for a theory does not necessarily apply to all of that theory, then it requires further argument to show that the empirical evidence we have for our scientific theories applies to the mathematics embedded in those theories. Why include the mathematical sentences of general relativity in the set of sentences taken to be confirmed by such empirical evidence as the observed advance in the perihelion of Mercury? It is tempting to say that we should include them because they are indispensable in deriving that phenomenon from the theory. But to say this is to slip back to an appeal to an account of evidential relevance along the lines of the H-D account and we have seen that this is problematic.

The indispensability argument is often stated as follows; we have good reason to believe our best confirmed scientific theories, mathematics is an indispensable part of those theories, therefore we have good reason to believe mathematics. Clearly however, we only have good reasons for believing the *well confirmed parts* of our scientific theories. We cannot simply assume without argument that empirical evidence applies to every part of a theory. If the mathematics used in a scientific theory is not confirmed by the evidence for that theory, then such evidence does not provide a good reason for believing that mathematics.

To argue that the mathematics used by a theory is confirmed or disconfirmed by empirical evidence therefore requires an account of evidential relevance which entails that empirical evidence is relevant to mathematics. The problems with the project of providing such an account are therefore problems which the indispensability argument cannot evade.

5. Eliminability and Field's Programme

We have seen that the mere fact that mathematical statements are included in an empirically confirmed scientific theory is not enough to show that the empirical evidence for the theory is also evidence for the mathematics included in it, since the evidence for a theory does not, in general, apply to the theory as a whole.

We can always add irrelevant hypotheses to a theory without impairing the ability of the theory to yield correct predictions, since classical logical consequence is preserved under addition of irrelevant assumptions. For instance we could add the hypothesis 'there are unicorns' to the general theory of relativity and it would still entail all the observations currently taken to confirm it. But these observations would not be taken to confirm the added irrelevant hypothesis. The hypothesis is irrelevant in the sense that the observations would still follow from the theory, even without the hypothesis. Of course many mathematical statements are not irrelevant to the derivation of correct predictions from our theories in this sense. Mathematics is not only included in our scientific theories, it is indispensable to them in the sense that there are correct predictions of the theory which would not follow without that mathematics. However, even this is not enough to show that such mathematics is confirmed by the predictions it entails since, as I argued in section

three, this sort of indispensability is not sufficient for confirmation by the entailed prediction.

It might be suggested at this point that there is another sense in which mathematics is not merely included in scientific theories, but is also indispensable to those theories. Mathematics is used in science, not only in deriving predictions from theories, but also in formulating the empirical hypotheses of those theories. Mathematical language enters into our descriptions of the physical world in a very fundamental way; there a fundamental features of the world which we cannot even state without reference to mathematical objects. Kepler's third law would be an example. This law, if it is true, commits not only to the existence of *planets*, but also to the existence of *numbers* – in particular, to the real numbers which are the values of the functions T(p) and D(p). The law states that there is a certain mathematical relationship between the values of these functions. Of course, this example is quite representative. Ever since Pythagoras first noted the mathematical relationships involved in the harmonies generated by vibrating strings, scientists have made use of the language of mathematics to formulate general truths about the physical world. This use of mathematics to describe the physical world has of course, been immensely successful. It seems that we cannot describe the structure of the physical world without presupposing a great deal of mathematics. This is a point that has been continually stressed by Hilary Putnam. In 'What Is Mathematical Truth?' for example, he writes:

> …[c]onsider a physical law, e.g Newton's Law of Universal Gravitation. To say that this law is true… one has to quantify over such non-nominalistic entities as forces, masses, distances. Moreover…to account for what is usually called 'measurement' – that is, for the numericalization of forces, masses and distances – one has to quantify not just over forces, masses and distances construed as physical properties…but also over *functions from* masses, distances etc. *to* real numbers….if one is a realist

> about the physical world then one wants to say that the Law of Universal Gravitation makes an objective statement about bodies – not just about sense data or meter readings. What is that statement? It is just that bodies behave in such a way that the quotient of two numbers *associated* with the bodies is equal to a third number *associated* with the bodies. But how can such a statement have any objective content at all if numbers and 'associations' (i.e functions) are alike mere fictions?
>
> [Putnam 1975c, p. 74]

Of course it is not just that we need mathematics in order to state physical laws like Kepler's Third Law or Newton's Law of Gravitation. We also need mathematics in order to state auxiliary hypotheses. We also want to say things like 'the orbital period of Mars is 1.87 years' or 'the mass of the electron is 9.10×10^{-31} kg'. And of course, it would be impossible to properly formulate our physical theories without a great deal of pure mathematics.[21]

Mathematics is used in science then in at least two different ways. We use mathematics *inferentially*, to derive predictions from theories, but we also use it *descriptively*, to make assertions about the physical world. Hence mathematics might be indispensable in science in two ways; it might be indispensable in deriving the observations which test our theories and it might be might indispensable in the very formulation of those theories. In this second sense, mathematics is indispensable in scientific theories in that it is impossible to describe certain physical phenomena without making use of mathematics. That is, there are no alternative theories of the same phenomena which do not refer to mathematical objects; reference to mathematical objects is ineliminable from scientific discourse.

This suggests another version of the indispensability argument, one which does not appeal to the indispensability of mathematics in deriving predictions, but appeals instead to

[21] See also [Putnam 1971, pp.35-43].

the *ineliminability* of statements committed to the existence of mathematical objects in the formulation of scientific theories. The argument would be that since reference to mathematical objects cannot be eliminated from our best scientific theories, then we should accept the mathematics used by those theories. Putnam states the argument succinctly, 'quantification over mathematical entities is indispensable for science...therefore we should accept such quantification; but this commits us to accepting the existence of the mathematical entities in question' [Putnam 1971, p. 57. See also Quine 1957, 1958].

In *Science Without Numbers* [Field 1980] Hartry Field attempted to undermine this version of the indispensability argument by showing that mathematics can be eliminated from science. Field develops a sophisticated *fictionalist* account of mathematics. He accepts the premise of the first horn of Benacerraf's dilemma, that mathematical discourse is irredeemably committed to the existence of mathematical objects and that these objects are abstract. Hence, if mathematics is true, then the 'standard account of truth' forces us to accept the existence of abstract mathematical objects. But Field denies that mathematics is true. His fundamental philosophical position is *nominalism*, the thesis that there are no abstract objects. Since mathematical statements commit us to the existence of various abstract objects, mathematics according to Field, is literally *false.*

But why should we accept the nominalist thesis? Field here appeals to the problems with abstract objects associated with the second horn of Benacerraf's dilemma; the problem of explaining how we could know anything about objects which exist nowhere in space-time and exert no causal influence on the world. What possible reason could there be for thinking that there really are any objects of this kind? In fact, Field thinks there could be only one good reason for believing in abstract mathematical objects, that we need to postulate them in our scientific theories of the physical world. Hence he aims to advance the cause of nominalism, by seeking to undercut what he takes to be 'the only non-question begging argument' for the existence of mathematical objects, namely the indispensability argument.

Field's argument is based on two claims. First, he claims that mathematics is dispensable in science, in the sense that attractive nominalistic versions of our scientific theories can be constructed. Let MP be some physical theory, formulated in mathematical language. Let α be any nominalistic assertion, that is, any assertion which does not quantify over mathematical objects. The aim of Field's programme is to show that for each such theory MP (that we actually accept) we can construct a *nominalistic* theory NP with the following property:

(N) $$MP \vdash \alpha \leftrightarrow NP \vdash \alpha$$

That is, the theory NP has all the same nominalistic consequences as MP. This means that reference to mathematical objects is eliminable from the theory MP, in the sense that there is an empirically equivalent theory which does not quantify over any mathematical objects.

Now Field does not seek to establish the very general claim that for *any* theory such as MP, there is *always* a nominalistic theory NP which satisfies (N). Instead, he argues for the dispensability of mathematics in formulating physical theories by appeal to what he hopes is a representative example; he constructs a nominalistic version of Newtonian physics (NNP) and shows that it satisfies (N). He further suggests that the techniques he uses to nominalize Newtonian physics could be extended to yield nominalistic versions of other physical theories. Hence, he attempts no general argument for the dispensability of mathematics in science; we have to look at our theories case by case and see if nominalistic

versions of them can be constructed. By showing how a representative theory can be successfully nominalized, Field hopes to show that there are reasons for some optimism about the success of such a programme of nominalizing scientific theories.[22]

If Field's programme were to succeed, then the ineliminability version of the indispensability argument would appear to be undermined, since he would have shown that, contrary to the main premise of that argument, our physical theories can be formulated without the use of any mathematics at all. Field goes further however. Even though mathematics is not *required* for the formulation of scientific theories, he wants to explain why mathematics may nonetheless be very *useful* in science. Here Field appeals to the inferential use of mathematics; mathematics is useful in science because it provides us with powerful techniques for deriving the empirical consequences of our theories. Field argues however, that we can explain the utility of mathematics in deriving empirical predictions from theories without assuming that mathematics so used is *true*. Instead, we can explain the reliability of mathematics in this respect in terms of the *conservativeness* of mathematics over nominalistic theories. Let N be a nominalistic theory and α any statement expressible in the language of N. Suppose we add a mathematical theory M to N, yielding the combined theory N+M. Field's second claim is that N+M is a *conservative extension* of N, which is to say that if N+M entails α, then α is also entailed by N alone. In *Science Without Numbers*, Field established that this claim holds for the most general mathematical theory we have, namely set theory, by proving the following conservativeness theorem:

(C) $$N + ZF^{*} \vdash \alpha \rightarrow N \vdash \alpha$$

[22] See [Field 1989, p.129].

The theory ZF* is essentially the usual Zermelo-Fraenkel axiomatisation of set theory, but with some modifications. As it is usually formulated ZF set theory quantifies only over sets. It has no nominalistic vocabulary at all and therefore no nominalistic assertions can be deduced from it. Hence adding ZF to a nominalistic theory N will trivially yield a conservative extension of N. But it is essential to the use of mathematical theories as tools for deriving the empirical consequences of physical theories that there is a connection between the nominalistic assertions of the theory and the mathematical assertions used to derive their consequences. To get a useful conservativeness result, Field augments ZF so as that it can make use of the nominalistic vocabulary of N.[23] The resulting theory is ZF*. The result Field proves – (C) above - is that adding ZF* to any body of nominalistic assertions N, yields a conservative extension of N; any nominalistic assertion which is a consequence of N+ZF* is also a consequence of N alone.

Now the conservativeness of a mathematical theory certainly does not imply that the theory is true. In fact, it implies little more than that the theory is consistent. This opens the way for an explanation of the inferential utility of mathematics in science which does not commit us to the truth of mathematics. According to Field, what is true in our physical theories is captured entirely by a nominalistic reformulation of the theory. However, the statements involving only nominalistic vocabulary which capture the true content of our theories will in general be represented by formulas which can be incredibly long and

[23] The augmentation proceeds as follows. Firstly, we allow our set theory to include so called *urelements*, elements of sets which are not themselves sets. Call the resulting theory ZFU. To ensure a sharp separation between the mathematical vocabulary of ZFU and the non-mathematical vocabulary of N, we add to ZFU a new predicate, Mx (for '*x* is a mathematical object') governed by three axioms; (a) everything that has an element is a mathematical object (b) the empty set is a mathematical object and (3) there is a set of all non-mathematical objects. We then relativise all the quantifiers of N to the non-mathematical objects, ~M. Finally, we allow the vocabulary of N to appear in the separation axiom of ZFU, ensuring that for any property definable in the language of N, there is a set of all non-mathematical objects with that property.

complicated. Correspondingly, showing that such a statement has a certain empirical consequence will, in general, involve a proof that is also long and complicated.

This is where mathematics comes in useful. Adding an appropriate mathematical theory, like ZF* to our nominalistic theory enables us to prove *mathematical counterparts* of the nominalistic assertions. We can then apply all the power of the mathematical theory to extract the nominalistic consequences of those assertions. But by (C), we know that any statement we derive in this way can already be derived from the nominalistic theory alone, although perhaps in a much more long-winded fashion. Hence, although adding mathematics to a nominalistic theory will not enable us to prove any new nominalistic statements, it may make the derivations of such statements more concise or systematic. Furthermore, (C) shows that it is always *safe* to add mathematics to our theories of the physical world, since doing so can never lead to any false assertions about the physical world which are not already implied by those theories.

On this view mathematics is simply a useful, but in principle dispensable tool for establishing the empirical consequences of our theories. But we can explain the reliability of this use of mathematics in science without supposing that mathematics is true. All we need to suppose is that mathematics is conservative and conservativeness does not imply truth.[24]

To see how this works, let us look in more detail at the strategy Field develops for constructing nominalistic reformulations of physical theories. In fact, there is a very simple

[24] Notice the formal analogy between Field's programme and Hilbert's formalist programme, discussed in chapter one. For both philosophers, there is a core set of statements (finitary statements in Hilbert's case, nominalistic statements in Field's case) which are epistemologically unproblematic. For both, classical mathematics is given an instrumental justification; we can use mathematics to more easily prove statements in the core. In both cases, it is argued that extending the core by the addition of a mathematical theory yields a conservative extension of the core. Hence we can explain the utility and reliability of the extended theory without supposing that it is true.

way in which Field could have argued for the thesis (N) above. Suppose we have a mathematically formulated physical theory MP, which combines mathematical and non-mathematical vocabulary. Suppose we can sort the predicates of MP into two kinds; those that apply only to mathematical objects and those that do not. Let MP_N be the set of all sentences of MP which make use only of the non-mathematical vocabulary. MP_N is then a theory which captures the entire nominalistic content of MP. Furthermore, by Craig's Theorem, MP_N can be recursively axiomatized.[25] That is we can construct a set of axioms in purely nominalistic language, from which we can deduce every nominalistic consequence of the mathematically formulated theory MP. In this way, we can establish that (N) holds in full generality, in the sense that for *any* mathematical physical theory MP, there is a axiomatisable nominalistic theory N which entails every nominalistic sentence entailed by MP.

Field is well aware of this result of course, but rejects it as a strategy for constructing nominalistic versions of physical theories on the grounds that it is 'trivial' and the resulting theories 'uninteresting' and 'unnatural' [Field 1989 p. 129,133]. He writes that such theories are 'obviously uninteresting since they do nothing whatever toward explaining the phenomena in question in terms of a small number of principles'. [Field 1980, p. 41, see also Hellman 1989, p. 135]. Field's aim is not simply to enumerate the nominalistic consequences of our mathematical-physical theories, but to construct *attractive* and *explanatory* nominalistic versions of those theories. [Field 1980, p. 47].

The strategy Field adopts is modelled on Hilbert's axiomatisation of geometry. The basic idea is to reformulate physical theories which make use of numerical functions (like

[25] Provided only that the MP can be given a first-order formulation and that the axioms of MP can be recursively enumerated. See [Craig 1953,1956]. See also [Putnam 1965] for an excellent discussion of the philosophical significance of Craig's result.

T(p) and D(p)) in terms of certain comparative, relational predicates. Let us consider just the geometrical portion of Field's nominalistic version of Newtonian physics, NNP. The *metric* approach to geometry begins by assigning a co-ordinate system; a function which maps each point of space to a unique triple of real numbers. Then we can introduce a *distance function* $d(p_1, p_2)$ which assigns a real number to each pair of points $<p_1, p_2>$, representing the distance between p_1 and p_2 in the given co-ordinate system. We can then define lines, curves and shapes as sets of points whose co-ordinates satisfy some algebraic expression, $y = mx + c$ for the straight line, $r^2 = x^2 + y^2$ for the circle and so on. In this way, we can give algebraic proofs of geometrical theorems.

The *synthetic* approach to geometry by contrast, dispenses with all these real numbers and functions. After all, geometry seems to be fundamentally concerned with points and lines and the relations between them, rather than with numbers. We can use co-ordinates and distance functions as a convenient way of *representing* geometrical facts, but the possibility of this kind of representation ought to be explicable in terms of the fundamental structural features of space, facts which we ought to be able to state without the use of the numerical representation. In fact, all we need are two primitive relations which can hold between points, the relations of *betweenness* and *congruence*. We can express these with a three-place predicate $x\mathrm{B}yz$ and a four-place predicate $xy\mathrm{C}zw$. These relations are taken as primitive and undefined, but intuitively we can think of $x\mathrm{B}yz$ as meaning that the point x lies on the line segment with end points y and z and $xy\mathrm{C}zw$ as meaning that the line segment with end points x and y is congruent to ('the same length as') the line segment with end points z and w. The axioms and theorems of Euclidean geometry can then be formulated in terms of these predicates. For example, Euclid's second axiom, that any line segment can be indefinitely extended can be expressed as $\forall x \forall y \exists z(yBxz)$.[26]

In this way, the nominalist can give an axiomatisation of geometry which dispenses with all reference to abstract objects like real numbers and functions and refers only to points of space and various relations which hold between them. Let NG (for 'nominalistic geometry') be the theory obtained in this way. We can then prove the following, so called *representation theorem*:

(R) For any model of NG, there is a function d which maps pairs of points in the model into the real numbers and which satisfies the following conditions:

(i) $\forall xyz$: $xyCzw \leftrightarrow d(x, y) = d(z, w)$

(ii) $\forall xyzw$: $xByz \leftrightarrow d(y, x) + d(x, z) = d(y, z)$

That is, the function d corresponds exactly to the intuitive interpretation of the predicates $xyCzw$ and $xByz$. Suppose we add an appropriate mathematical theory, like ZF*, to NG. Then the representation theorem entails that we can define a function d such that the equivalences in (i) and (ii) hold in any model of the combined theory NG+ZF*. This function d allows us to obtain mathematical counterparts of the nominalistic assertions of NG. We can simply replace $xyCzw$ wherever it occurs in a formula with the right hand side of the equivalence in (i) and we can replace $xByz$ wherever it occurs in a formula with the right hand side of the equivalence in (ii)[27]. We can then use the techniques made available by the mathematical component of the combined theory to derive the consequences of our

[26] See [Hilbert 1971] for further details.
[27] So for example, the mathematical counterpart of the axiom $\forall x \forall y \exists z(yBxz)$ is $\forall x \forall y \exists z(d(x, y) + d(y, z) = d(x, z))$.

nominalistic assertions. But by the conservativeness theorem (C), we know that any nominalistic assertion derivable in this way already follows from the nominalistic theory NG alone.

As well as the representation theorem (R), we can also prove the following *uniqueness theorem*:

(U) The function $d(x, y)$ which satisfies conditions (i) and (ii) above is unique up to a multiplicative constant.[28]

Hence, the fact that geometric laws when formulated in terms of a distance function are invariant under multiplication of all distances by a positive constant, but are not invariant under other transformations, receives a satisfying explanation. It is explained by intrinsic facts about physical space, facts which can be expressed in the purely nominalistic language of NG.

In *Science Without Numbers*, Field showed how this strategy can be extended give a nominalistic formulation of classical metric field theories in flat (Euclidean) space-time. A *field* is usually represented as a function which assigns certain mathematical objects (numbers or vectors) to points of space-time. Field draws on work in measurement theory to show how reference to such functions can be eliminated in favour of comparative predicates of various kinds. He then shows how physical laws, stated in terms of such functions (such as Newton's Law of Gravitation) can also be restated in terms of those comparative predicates. The result is his nominalistic formulation of Newtonian physics,

[28] That is, if d_1 and d_2 are any two functions satisfying (i) and (ii) then d_1 and d_2 differ only by a positive multiplicative constant and conversely if d_1 and d_2 differ only by a positive multiplicative constant, then d_1 satisfies (i) and (ii) only if d_2 does.

NNP. Field sketches a proof of an appropriate representation theorem for NNP, showing how, if we add ZF* to NNP we can reintroduce the numerical functions used in the usual formulation of the theory and in this way obtain a mathematical representation of the underlying nominalistic theory. But the possibility of introducing such a representation is regarded as a derivative fact, which is explained in terms of the intrinsic features of space-time captured by the nominalistic theory.

If MNP is the usual mathematical formulation of Newtonian physics, then the representation theorem implies that, for any nominalistic assertion α:

(1) $\text{MNP} \vdash \alpha \leftrightarrow \text{NNP} + \text{ZF}^* \vdash \alpha$

But by the conservativeness of mathematics (C), we have:

(2) $\text{NNP} + \text{ZF}^* \vdash \alpha \rightarrow \text{NNP} \vdash \alpha$

Since NNP is a proper subset of NNP+ZF*, we also have:

(3) $\text{NNP} \vdash \alpha \rightarrow \text{NNP} + \text{ZF}^* \vdash \alpha$

Hence:

(3) $\text{NNP} + \text{ZF}^* \vdash \alpha \leftrightarrow \text{NNP} \vdash \alpha$

Combining (1) and (3) we get:

(4) $\text{MNP} \vdash \alpha \leftrightarrow \text{NNP} \vdash \alpha$

and Field has succeeded in proving an interesting and representative instance of (N) above.

Field argues that synthetic formulations of physical theories are more illuminating than the usual metric formulations, not only because they are nominalistic but also because they explain the phenomena without appeal to *extrinsic*, causally irrelevant objects. The mathematical objects referred to in the metric formulation of a theory play no causal role in explaining the physical facts. They are invoked as something extrinsic to the process to be

explained, an entity related to that process only by an arbitrarily chosen function. Field's nominalistic formulations of physical theories show how we can give a purely *intrinsic* explanation of the phenomena, an explanation that does not invoke arbitrary and causally irrelevant entities:

> I am saying then that not only is it much likelier that we can eliminate numbers from science than electrons (since numbers, unlike electrons, do not enter causally in explanations), but also that it is more illuminating to do so. It is more illuminating because the elimination of numbers, unlike the elimination of electrons, helps us to further a plausible methodological principle: the principle that underlying every good extrinsic explanation there is an intrinsic explanation.
>
> [Field 1980, pp. 44-5]

If Field's programme of nominalizing physical theories in this way could be carried out for every theory we currently accept, this would show that mathematics *is* dispensable in scientific theories in the very strong sense that there are *better* theories of the same phenomena which make no use of mathematics at all. On the other hand we can explain why adding mathematics to our best nominalistic theories is useful in drawing out the nominalistic consequences of our theories. even though that mathematics is literally false. For by the conservativeness of mathematics, we can always be sure that adding mathematics to our theories will never enable us to derive any new nominalistic conclusions which could turn out to be false.

Both aspects of Field's project have been criticized, however. Consider first Field's explanation of the inferential utility of mathematics in science. Stewart Shapiro has pointed out that Field's conservativeness claim (C) can be interpreted in two ways, depending on whether we interpret the relation of logical consequence in terms of deductive or semantic entailment. The proof of (C) given by Field establishes only the semantic conservativeness of mathematics. In a second-order language such as the language Field uses to formulate NNP, deductive and semantic entailment do not coincide, so we cannot infer from the semantic conservativeness of mathematics that it is also deductively conservative. In fact, Shapiro shows that if we interpret entailment deductively, the conservativeness claim fails for NNP. In more detail, Shapiro proved that for any mathematical theory M, powerful enough to prove the consistency of Field's second-order nominalistic theory NNP, there are nominalistic statements which are deductive consequences of NNP+M but which are not deductive consequences of NNP alone. That is, NNP+M is not a deductively conservative extension of NNP. It is however, a semantically conservative extension of NNP; every nominalistic statement which is a semantic consequence of NNP+M is also a semantic consequence of NNP. Shapiro argues however that Field needs to appeal the deductive conservativeness of NNP+M, since he wants to explain the utility of mathematics in science in terms of its power to shorten *deductions*.[29]

There has also been much debate concerning the adequacy of Field's strategy for constructing nominalistic reformulations of scientific theories. Firstly, it has been argued that Fields' version of Newtonian physics NNP is not a nominalistically acceptable theory. For NNP quantifies not only over space-time points, but also over *regions* of space-time points. It makes use of a powerful second-order logic, which appears to give us a thinly disguised way of quantifying over *sets* of physical objects. Secondly, even if NNP is nominalistically acceptable, it is not clear how Field's techniques could be extended to

[29] See [Shapiro 1983a]. For Field's response to this problem see [Field 1985,1989 pp. 125-146].

more complex theories such as quantum mechanics in any nominalistically acceptable way.[30]

Irrespective of whether Field's reformulations of physical theories are *nominalistically* adequate, we can also ask whether they are *scientifically* adequate. For example, it has been argued that Field's nominalistic language is inadequate to the needs of science in terms of its expressive power. That is, there are statements scientists are interested in establishing (statements about certain properties of field theories for example) which cannot be expressed in Field's nominalistic language.[31]

Even if these problems could be solved, it is not clear that the success of Field's programme would fatally undermine the indispensability argument. Suppose Field did succeed in giving an adequate nominalistic reformulation of Newtonian physics. We would then have *two* theories to consider, Field's nominalistic Newtonian physics NNP and the usual mathematically formulated version MNP. The question is then which is the *better* theory.

One way in which Field argues that his nominalistic reformulations are better than their mathematical counterparts is by appeal to the principle of parsimony or Occam's razor; given two theories of the same phenomena, *other things being equal*, we should accept the theory with the fewest ontological commitments.[32] This argument cannot be

[30] See for example, [Malament 1982, p. 532-4, Hale 1988, Shapiro 1983a]. Field anticipates and responds to some of these objections in *Science Without Numbers*. See also [Field 1985, 1989, pp. 171-281].
[31] See [Malament 1982, p. 528, Melia 1998, p. 69].
[32] Just how we are to measure the ontological commitments of theories in order to make such comparisons is not obvious however. In terms of the number of individuals postulated, NNP and MNP are on a par. Both are committed to uncountably many objects, uncountably many space-time points in NNP, uncountably many real numbers in MNP. In terms of the number of *types* of object postulated however, NNP has fewer commitments, since MNP is committed to at least one type of object that Field's version is not, namely infinitely many abstract mathematical objects. On the other hand, if we formulate MNP so that it dispenses with space-time points by simply identifying them with their co-ordinates, then we can say the same thing

conclusive however. Occam's razor is only one consideration among many which become operative when trying to decide between otherwise equivalent theories. In addition, considerations of simplicity, elegance, explanatory power and coherence with other theories may become operative. There are no doubt many other types of consideration active in such decisions and they do not all have equal weight. There is no reason for thinking that ontological parsimony is an over-riding factor in such decisions. I doubt that working scientists would find Field's nominalistic version of Newtonian physics *more* acceptable than the usual mathematical versions, merely on the grounds that it has a reduced ontology.

I think that Field recognises this point and that is why he appeals to considerations other than ontological parsimony. For, as already mentioned, according to Field other things are *not* equal. He argues that the problem with the usual metric formulations of field theories is that invoke arbitrary, causally irrelevant objects, extrinsic to the phenomena to be explained. Field's synthetic formulations of those theories, by contrast, dispense with causally irrelevant objects and explain the phenomena in terms of intrinsic features of reality in such a way that we can explain why the metric formulations are arbitrary to exactly the extent that they are. Hence, according to Field, the synthetic approach has many advantages quite apart from considerations of ontological parsimony or nominalistic scruples about abstract objects.

The case Field presents here is far from clear-cut however. It is not clear that the synthetic formulations of theories really do any better in this respect than the metric theories they are intended to replace. Joseph Melia, for example, has argued that in fact,

about NNP; it is committed to at least one kind of object (namely space-time points) that MNP is not. See also [Melia 1998, pp. 67-68] for further doubts about the ontological parsimony of Field's nominalistic theories.

Field's reformulations also introduce entities which are extrinsic to the process to be explained, causally irrelevant and arbitrary.[33]

These sorts of considerations cannot settle the matter however. In fact, it is a mistake to formulate the indispensability argument in terms of the ineliminability of reference to mathematical objects in scientific theories. The more fundamental question concerns not the *eliminability* of mathematics from science, but the relevance of empirical *evidence* to mathematics.

It is important here to distinguish carefully between two different kinds of argument that can be made concerning scientific theories or hypotheses, arguments which operate at different methodological levels. At the first level, the level of *evidential relevance*, there are arguments which aim to show how some body of evidence is relevant to a particular hypothesis or theory. For example, Einstein argued that the observed facts concerning the photoelectric effect were evidence for the hypothesis proposed earlier by Planck, that electromagnetic energy is emitted and absorbed in discrete units or quanta.[34] Likewise Newton, by showing how the law of universal gravitation entailed Kepler's laws was arguing that the observed facts of planetary motion were relevant supporting evidence for his theory. In these cases, the evidence was confirming evidence. On the other hand, there are arguments which aim to show that some observation is relevant disconfirming evidence for a hypothesis. So for example, Michelson and Morley gave an argument aimed at showing that the results of their famous experiment provided evidence which disconfirmed the hypothesis that light waves propagate through the medium of the ether.

[33] See [Melia 1998].
[34] See [Gamow 1996, pp. 22-27].

That a theory is confirmed by some evidence is a necessary but not sufficient condition for it to be rational to accept that theory. We may also require that there should be no disconfirming evidence. Nor is only one confirming observation sufficient, a theory must be confirmed by a wide and varied body of evidence before it becomes rational to accept it.[35] The second level, the level of *theory choice*, is where we find arguments for accepting or rejecting a theory based on considerations of this kind. So for example, we might argue that we should accept Newton's theory of universal gravitation on the grounds that there is a wide range of supporting evidence for it; the orbits of the planets, the precession of the equinoxes, the motions of the tides and so on. We might also mention the fact that there is no competing theory that explains all these things as successfully. This would be a second level argument. It *mentions* the fact that there is some confirming evidence for Newton's theory, but it does not by itself *show* that the evidence mentioned is in fact relevant supporting evidence for the theory. The latter claim is provided by Newton's actual derivation of the planetary orbits, the motions of the tides and so on from his theory. It is at the second level where we also find arguments concerned to adjudicate between rival theories of the same phenomenon. So for example, we might compare the phlogiston and oxygen theories of combustion, or the classical and quantum theories of electromagnetic radiation. In debates like this, it is taken for granted that there is *some* evidence for both theories, otherwise the debate would be pointless. What we do is weigh

[35] This is somewhat simplified of course. It may be rational to accept a theory for which there is *some* disconfirming evidence, if there is a great deal of confirming evidence for it. What we in fact do is weigh up the successes of a theory against any failings it may have, where 'success' and 'failure' are measured in complex ways. We saw this process at work in the case of mathematics in our discussion of Kitcher's account of the rigorization of the calculus. My point here is just that a first level argument, to the effect that there is *some* evidence for a theory is required before any further second level questions come into play.

up the evidence for and against each theory and try to make a decision about which theory is the more acceptable.

Now it may happen that we have two theories of the same phenomenon which are empirically equivalent. All the empirical evidence we have for one theory applies equally to the other, the theories agree on all the observational data. The Ptolemaic and Copernican systems of celestial mechanics might be one example. Another would be the two equivalent versions of quantum mechanics, Schrödinger's wave equation formulation and the Heisenburg matrix representation. In such a case we may appeal to such considerations as ontological parsimony, elegance or simplicity, ease of use, explanatory power and so forth - the so called 'theoretical virtues'. Such arguments are also part of the second level.[36] Again, in such a debate, we are *assuming* that there is some relevant confirming evidence for each theory (in fact, in a case like this, we are assuming that all the relevant observational evidence for one theory is equally relevant confirming evidence for the other). But it requires a first level argument to *show* that there *is* some relevant confirming evidence for our theories.

The point of this distinction is just that we require an argument of the first kind if there is to be any argument of the second kind. Before we can choose between two competing theories, by weighing up the evidence for and against them, it is first necessary to show that there is some relevant supporting evidence for either theory.

The eliminability version of the indispensability argument appeals to considerations operative at the second level, the level of theory choice. As stated above, the argument was that since mathematics is ineliminable from scientific theories, in the sense that there is no

[36] Notice that second level considerations like these may decide in favour of one theory rather than another, as in the case of the Ptolemaic and Copernican systems, or they may not, as in the case of the two equivalent representations of quantum mechanics.

way of formulating those theories nominalistically, we should accept the mathematics used in those theories. This is a second level argument; we should accept a certain theory because there is no better theory of the same phenomena. But presumably we should accept our theories because they are well confirmed by evidence; confirmation is a necessary though not sufficient condition for accepting a theory. If a theory is not confirmed by some body of evidence, then second level arguments concerning the relative merits of two theories are irrelevant.

The ineliminability version of the indispensability argument is premised on the claim that mathematics is required in order to formulate our physical theories. Field attempts to show that this premise is false by constructing nominalistic version of those theories. This way of stating the indispensability argument makes the question turn on whether or not we can eliminate reference to mathematical objects from our physical theories. But the important question is not whether reference to certain objects is eliminable or not, but whether there is any evidence for those objects. Craig's theorem shows that so long as we can clearly distinguish between entities of one kind and those of another kind, reference to objects of either kind is eliminable in favour of reference to the other. So we can eliminate talk of electrons, photons and so on in favour of talk about meter readings and other observations. Likewise, we can eliminate talk of numbers and functions in favour of talk about space-time points, electrons and photons. But whether or not it would be a good idea to eliminate talk of electrons and photons depends on whether or not we have good reason to believe that they exist. That is, it depends on whether or not we have any evidence for their existence. If we do, then the fact that reference to them is eliminable is irrelevant.

On the other hand, even if reference to Xs is ineliminable from a theory, it does not follow that there is any evidence for the Xs. We can easily imagine any number of physical theories for which there is no evidence at all, but which could not be stated without referring to certain objects. Hence, the *mere fact* that a theory quantifies over mathematical objects does not show that there is any evidence for that theory and *a fortiori*, does not show that there is any evidence for the mathematics used in the theory.[37]

Of course, we do have plenty of scientific theories, for which there is a great deal of evidence and it may be that we cannot reformulate those theories nominalistically. But as argued in section four, the fact that there is some evidence for a theory does not show that there is evidence for the mathematics included in that theory, since the evidence for a theory may not be relevant to *every* statement included in the theory. This clearly entails that we should accept only those parts of our theories which are confirmed by the evidence we have for them. If none of that evidence confirms the mathematics included a theory, then no second level argument (to the effect that there is no way of stating the theory without referring to mathematical objects) can provide a good reason for accepting such mathematics.[38]

For the indispensability argument to work, we need to appeal to some holistic principle which entails that the evidence we have for our theories applies to the

[37] See also [Melia 1998, pp. 70-1].

[38] Notice however that on any adequate account of evidential relevance, there is *some* empirical evidence for mathematically formulated empirical hypotheses like Kepler's third law. Any evidence for such laws would seem also to be evidence for the mathematical objects those laws are committed to. However, it is not obvious how this argument can show that there is empirical evidence for any *pure* mathematical statement. Of course, in order to *use* Kepler's law, we need to make use of many statements of pure mathematics. In this sense Kepler's law 'presupposes' a great deal of pure mathematics; if not the full theory of the real numbers, then at least a theory of the rationals [see Putnam 1971, p. 55]. But what is needed here is an argument which shows how the pure mathematics presupposed by Kepler's law in this sense is confirmed by the very same evidence as confirms that law itself. I am not suggesting no such first level argument could be given, only that such an argument is required.

mathematics included in those theories. The ineliminability version of the argument does not by itself provide any such argument, it simply assumes some form of holism. I have argued that a certain extreme form of holism is false; we cannot assume that the evidence we have for a theory is evidence for every part of that theory. How could a more conservative holistic principle be argued for? We have seen that a common argument in favour of holism, the argument from the H-D account of confirmation, will not work. Any argument for holism will have to appeal to some other characterization of evidential relevance.

6. EXPLANATION

The first version of the indispensability argument we considered fails because the hypothetico-deductive account of confirmation on which it is based is inadequate. That a theory entails a true prediction is not sufficient for that prediction to confirm the theory. But this is puzzling. After all, we do test our scientific theories by deriving testable predictions from them. If entailment alone is not sufficient for confirmation, what extra condition is required? One answer is that the derivation of the prediction from the theory must provide an *explanation* of the predicted phenomenon. It is well known that entailment is not sufficient for explanation.[39] Perhaps what is required in addition to entailment for confirmation is exactly what is required in addition to entailment for explanation.

W.D Hart, in a footnote to 'Access and Inference' considers this idea:

[39] See for example [Salmon 1989, p. 47].

> I doubt that the notions of confirmation and explanation are independent. It is well known that triviality ensues from the thesis that what entails an observation is confirmed by it...entailment is not sufficient for confirmation. More likely, the best explanation of an observation is confirmed by it...If some mathematics is confirmed by observation, then that is because it is inextricably part of what best explains such observation.
>
> [Hart 1996 p. 56-57]

Suppose we replace the hypothetico-deductive account of confirmation in the indispensability argument with an *explanation criterion* for confirmation: if a conjunction of statements explain an observation, then the observation confirms the conjunction of hypotheses.[40] The corresponding indispensability claim would then be that there are some mathematical statements which play an indispensable role in the scientific explanation of observations. The indispensability argument will then go through. An observation confirms the conjunction of hypotheses required to explain it. By the distribution principle each of the conjuncts is also confirmed. By the indispensability thesis, some of those conjuncts will be mathematical statements. Thus mathematical statements can be confirmed (and by an analogous argument, disconfirmed) by observations.

This version of the indispensability argument appeals then to the principle of *inference to the best explanation*. Just as we need to postulate objects such as electrons and photons in order to explain our observations, we need to postulate the existence of a multitude of mathematical objects in order to explain the same observations. Hence the evidence we have for theories which postulate electrons and photons applies equally to the mathematical objects referred to in those theories.

Indispensability arguments are exactly as successful as the accounts of the relation of evidence to theory on which they are based. The explanation criterion for confirmation is

[40] Notice that if entailment is not *necessary* for explanation (as well as being insufficient) then this criterion need not even imply that the conjunction must entail the prediction which confirms it.

certainly an improvement over the H-D account and the corresponding indispensability argument is therefore on firmer ground. The appeal to explanation as opposed to deduction avoids many of the problems of relevance which beset the hypothetico-deductive account, since irrelevant hypotheses, even if they entail an observation, do not explain it.

According to the explanation criterion of evidence, if a mathematical theory enables us to explain various empirical phenomena, then there is at least some empirical evidence for that mathematical theory. But this gives us only a *prima facie* case for realism concerning mathematics. Whether or not it is rational to accept a theory depends not only on their being some supporting evidence for the theory but also on various second level considerations; whether there is any disconfirming evidence for the theory, the scope and variety of the confirming evidence and the relative merits of any competing theories of the same phenomena.

From this perspective, the success of Field's attack on the indispensability argument depends on whether the explanations of the phenomena provided by his nominalistic reformulations of physical theories are better than the those provided by the mathematical formulations. If the nominalistic explanations are better, then inference to the *best* explanation will tend to support the nominalistic theories as opposed to the mathematical theories. Here Field could appeal to the principle of parsimony and to the advantages of intrinsic over extrinsic explanations. As we have seen however, the argument here is far from conclusive.

Nonetheless, the explanation criterion version of the indispensability argument does face some difficulties. On the role of mathematics in scientific explanations, Field writes:

> ... *even on the assumption that mathematical entities exist*, there is a *prima facie* oddity in thinking that they enter crucially into explanations of what is going on in the non-platonic realm of matter. ....the role of mathematical entities, in our explanations of the physical world, is very different from the role of physical entities in the same explanations. For the most part, the role of physical entities in those explanations is causal: they are assumed to be causal agents with a causal role in producing the phenomena to be explained. Since mathematical entities are assumed to be acausal, their explanatory role (or roles) must be somehow different.
>
> [Field 1989, pp.18-19]

Mathematical objects, if they are abstract, are causally isolated from the physical universe. It is usually thought to be a failure of explanation to invoke entities which are causally irrelevant to the phenomena to be explained. Perhaps this sort of causal asymmetry in the explanatory roles of physical and mathematical entities provides a reason for thinking that an observation explained by theory does not confirm the mathematical component of the theory. If causation is the missing factor which turns entailment into explanation, then abstract mathematics will not explain anything.[41]

Cheyne and Pigden put the point in terms of a dilemma for the platonist: either mathematics is dispensable in science, in which case there is no reason to believe there are any mathematical objects at all, or mathematics is indispensable in science, and hence mathematical objects must play a causal role in explaining what is going on in the physical world.[42] But if mathematical objects are causal agents, platonism is not the correct account of them. So in either case, platonism is refuted. This is too quick however. It is right to point out that the indispensability argument, properly understood, requires that mathematics plays a role in explaining physical phenomena, but as Mark Colyvan points out in his

[41] See also [Melia 1998, pp. 70-1].
[42] See [Cheyne and Pigden, 1996].

discussion of Cheyne and Pigden's paper, it is a mistake to think that any such explanation must be causal.[43]

Causal connection cannot be a necessary condition for explanation, since not all explanations are causal. Causation is fundamentally a relation between events rather than statements and it is not only the occurrence of some event that is capable of explanation. General laws can also be explained. For example, Newton was able to explain Kepler's third law by showing it was a consequence of his inverse square law of gravitation. In this way, the evidence for Kepler's third law accrued automatically to Newton's law. Yet there is no causal relation between the two laws.

Furthermore, mathematical truths themselves are capable of explanation. Some proofs of a theorem, show not only *that* the theorem is true, but also show *why* the theorem is true. Just what this means is the subject of chapter seven. For now, notice that in explaining a mathematical truth, we appeal to other mathematical truths and these are not taken to be causes of the mathematical fact being explained. A statement may be explanatorily relevant to another, without being in any way causally relevant.

The problem facing this version of the indispensability argument is again one of relevance. It requires an account of explanation which allows for mathematical statements to be explanatorily relevant to empirical phenomena. It is well known that it is no simple matter to give an adequate account of our concept of explaining something. Nonetheless, I do not think this represents an insurmountable problem for the indispensability argument. Since not all explanations are causal, it may well be possible to give an account of

[43] See [Colyvan 1998a].

explanation which shows how mathematics can play an explanatory role in science[44]. If such an account can be provided, then the indispensability argument will have shown that there is a great deal of empirical evidence for mathematics.

7. Empirical vs. Non-Empirical Evidence

The first version of the indispensability argument we considered attempted to show that the evidence which confirmed or disconfirmed a theory also confirmed or disconfirmed the mathematics required to entail that evidence. I argued that this presupposes a certain account of confirmation, namely the hypothetico-deductive account and that this account is inadequate.

We then considered a version of the indispensability argument which attempted to bypass the difficulties involved in providing an account of evidential relevance. The argument assumed that evidence always confirms the whole of a theory. I argued that this version of the argument is also unsound because the assumption is false. Evidence can be relevant to less than the whole of a theory.

I then considered a version of the argument which appealed to the ineliminability of mathematics in formalizing scientific theories. I argued that for this version of the argument to work, we would again need to appeal to some holistic principle that evidence confirms all of a theory, or a subset which typically includes some mathematics. Such a holistic principle could only be justified by appeal to some account of the confirmation of a theory

[44] Mark Colyvan in 'Confirmation Theory and Indispensability' suggests that perhaps the idea of *explanation as unification* will help here [Colyvan 1999]. In that paper, Colyvan places a similar emphasis on the role of theories of evidence in the analysis of the indispensability argument to that developed here and arrives at similar conclusions.

by evidence. We cannot just say 'quantification of mathematical objects is indispensable in scientific theories, therefore we ought to accept such objects'. We have to show that the evidence we have for our theories applies to the mathematics included in them. Once we have done that, we can go on to consider whether there are any better theories, which do not quantify over mathematical objects. But if we do not have an adequate account of the first level relation of evidence to theory, which succeeds in establishing that the evidence for a theory also applies to the mathematics included in the theory, then such second level arguments are irrelevant.

Finally, we looked at a version of the argument which appealed to an explanation criterion for confirmation. This version of the argument does not attempt to bypass the problem of stating the relation between theory and evidence and the account of evidential relevance it provides avoids many of the problems with the hypothetico-deductive account. However, for this version of the argument to be convincing, it must provide a satisfactory account of explanatory relevance.

Although there are problems with the indispensability argument, it would be a mistake to leap to the conclusion that there is no such thing as empirical evidence in mathematics. What I have been trying to show is that indispensability arguments run up against the severe difficulties involved in providing an account of the relation of evidence to theory. But clearly, the fact that there are problems in giving an adequate account of this relation, does not imply that there is no such thing as evidence for a theory. Likewise, the problem of describing the relationship of empirical evidence to mathematics (the problem as I see it, of giving an account of how mathematics can explain physical features of the world) does not show that there cannot be empirical evidence for mathematics.

In fact, to deny that there can be empirical evidence for mathematics would involve a conflict with mathematical practice. For mathematicians do sometimes appeal to the utility of mathematical theories in science as providing some evidence for those theories. An obvious example would be the calculus of the early seventeenth century. Although the rigour of the proofs of the fundamental principles of the new theory could be criticised, many mathematicians argued that the immense success of the theory in solving problems in physics showed that the theory must be getting something right. Leibniz, for example, argued that the success of the calculus made it rational to put aside worries about rigor and urged the mathematical community to find ways of extending the power of the calculus "because of the application one can make of [the calculus] to the operations of nature, which uses the infinite in everything it does." [Leibniz 1849-63, vol. 2, p. 219. See also Kitcher 1984, p. 236].

Another example concerns the theory of complex numbers.[45] It can be argued that *one* factor in the gradual acceptance of complex numbers was their utility in solving problems that cropped up in the sciences. For example, the complex numbers allowed for a great simplification in the theory of solutions to partial differential equations. In this way, complex numbers proved to be immensely useful in a wide and varied range of physical applications (heat-flow, fluid dynamics and so on) where such equations are common.[46]

---

[45] On the history of complex numbers see [Kline 1972. pp. 502-43,671-708]. On the utility of complex numbers in scientific applications see [Colyvan 1999].

[46] A mathematical statement need not be embedded in a scientific *theory* for there to be empirical evidence for it. Consider the theorem of analytic geometry, that for a curve given by a function $f(x)$ the length of the curve from a to b is given by the formula $\int_a^b \sqrt{1 + f'(x)^2}\, dx$. We can test this statement empirically by drawing a particular curve and then laying a piece of string along it from a to b, so that it follows the line of the curve. We can then straighten the string out, measure its length, and see if this agrees with the result given by the formula. Examples of this kind of empirical confirmation of a mathematical statement could multiplied endlessly.

On the other hand, mathematical theories seem highly resistant to disconfirmation by empirical evidence. Mathematicians do not in practice take the empirical disconfirmation of a scientific theory as applying to the mathematical component of the theory. Consider the case of Euclidean geometry. It is often suggested that the general theory of relativity has shown that Euclidean geometry is *false*.[47] If so, then we would have an empirical disconfirmation of a mathematical theory. This analysis is misleading however. What has happened is that Euclidean geometry has been *reinterpreted*.

For nearly two thousand years, mathematicians believed that Euclidean geometry was the only mathematically possible geometry. Since Euclidean geometry was thought of not only as a mathematical theory, but also as a theory of physical space, it was also held that physical space had to be Euclidean. Both of these beliefs have since been overthrown. Firstly, mathematicians discovered that it was possible to develop an alternative to Euclidean geometry based on the negation of the parallel postulate. This postulate is equivalent to the claim that given any line l and any point p not on that line, there is exactly one line through p that is parallel to l. Sacherri [1733] had attempted to prove the parallel postulate by *reductio ad absurdum*. He was able to derive a contradiction from the assumption that there is *no* line parallel to l through p, but was unable to derive a contradiction from the assumption that there is more than one such line. Bolyai [1832] and Lobachevsky [1855] independently developed the geometry (now known as *hyperbolic* geometry) in which there are an infinite number of parallels to any given line through a point not on that line. Bolyai and Lobachevsky were convinced that this geometry was consistent and that the question of its applicability to physical space could not be settled by

[47] See for example [Putnam 1994c].

a priori mathematical argument but only by experience.[48] Beltrami and Poincaré were later able to prove that the geometry of Bolyai and Lobachevsky is consistent relative to Euclidean geometry, by constructing models of it within Euclidean geometry itself.[49]

Meanwhile, Gauss was developing the field of differential geometry by studying the properties of curved surfaces in three dimensional (Euclidean) space.[50] This work revealed that the fundamental geometrical properties of a surface (such as the distance between two points connected by a curve on the surface and the curvature of the surface at a point) depend only on *intrinsic* properties of the surface. That is, they do not depend on the way the surface is embedded in three dimensional space. This implies that we can consider a surface to be a *space in itself*, since the geometrical properties of the surface are quite independent of the surrounding space. If we consider a surface as a space in its own right, and take the 'straight lines' of the surface to be *geodesics* (the curves on the surface which represent the shortest distance between two points) then the geometry of the space is non-Euclidean.[51]

Riemann further developed this implication of Gauss's work on surfaces. He introduced the concept of an arbitrary n-dimensional space (or *manifold*) and showed how to define a measure of curvature at each point in such a space. If the value of the curvature at a point is the same as that at every other point in the space, then we have a space of constant curvature. In such spaces, geometrical figures can be moved around without changing their shape. Furthermore, if the space has a constant curvature of *zero*, the geometry of the space is Euclidean. If the space has a constant *positive* curvature, then we

[48] See [Kline 1972, pp. 861-81].
[49] See [ibid. pp. 913-7].
[50] See [Gauss 1827]. Gauss also appears to have independently discovered the non-Euclidean geometry of Lobachevsky and Bolyai, but never published the results of his investigations. See [Kline 1972, p. 871-3].
[51] See also [Kline 1972, pp. 882-889].

obtain a new non-Euclidean geometry. In two dimensions, we obtain the intrinsic geometry of the surface of a sphere, sometimes called *spherical* or *double elliptical* geometry. The straight lines of this geometry are the great circles of the sphere. Hence in this geometry, all lines are of finite length and there are no parallel lines, since any two lines intersect at two points. On the other hand, if the space has a constant *negative* curvature, we obtain the non-Euclidean geometry of Bolyai and Lobachevsky in which there are an infinite number of parallels to any given line.[52]

These developments led quite naturally to the suggestion that geometrical theories in mathematics should be reinterpreted.[53] We should see them not as theories of physical space, but as theories of certain abstract spaces. Euclidean geometry is the theory of one kind of space (a space of constant zero curvature), hyperbolic and spherical geometry describe different kinds of space (spaces with constant negative or positive curvature respectively). The alternative geometries are not then competing accounts of physical space, but simply equally correct theories of distinct abstract spaces or structures. Under this new interpretation of geometry, the question which of these structures best describes physical space was no longer a mathematical question, but could be left for the physicists to determine.

From this perspective, what the success of general relativity shows is that Euclidean geometry is not true of physical space.[54] Mathematicians have not abandoned Euclidean geometry in the sense that physicists have abandoned the phlogiston theory of combustion.

[52] See [Riemann 1854]. See also [Kline 1972, pp. 889-6, 904-23, Le Corbeiller 1954].
[53] This suggestion was made most explicitly by Riemann in his paper 'On the hypotheses which lie at the foundations of geometry'. [Riemann 1854].
[54] In general relativity, gravitation is explained by supposing that the curvature of space *varies* from point to point. Hence neither the hyperbolic geometry of Lobachevsky and Bolyai, nor Riemann's spherical geometry of constant positive curvature are true of physical space either.

Research still goes on in Euclidean geometry; new theorems are proved in it and old theorems are given new proofs. Furthermore, these proofs are taken as proofs of *true* statements. It would be misleading to say that we now take the statement 'the angles of a triangle add up to 180 degrees' to be simply *false*. What we would now say is '*in Euclidean space*, the angles of a triangle sum to 180 degrees'. This is a *true* statement about a certain abstract structure, a structure which is just one member of the class of mathematical spaces.

I have argued that mathematical theories can be empirically confirmed by appeal to their utility in scientific theories and cited the examples of the calculus and complex numbers. But if mathematical theories are capable of empirical *confirmation*, they ought also be open to empirical *disconfirmation*. And yet, mathematical theories are never empirically disconfirmed by the failure of a scientific theory of which they are a part, Euclidean geometry is a case in point. This situation seems puzzling. How can we explain this asymmetry between the empirical confirmation and disconfirmation of mathematical theories?

Recall that Quine's solution to this problem is to appeal to pragmatic considerations. The reason we are less inclined to revise the mathematical components of our theories in the face of empirical disconfirmation of them is simply that doing so would involve us in a wide-ranging revision of the total system of science. A pragmatic desire to change our overall system of beliefs as little as possible then counsels us against revision of the mathematical components of our theories.

What this fails to take into account is the fact that we can have better evidence for some scientific hypotheses than others. As Putnam points out, the reason why we usually revise one or more of the auxiliary hypotheses used to derive a prediction from a hypothesis is that the auxiliary hypothesis are often far less well supported by evidence than the hypothesis under test.[55] Indeed, sometimes the auxiliary hypotheses are even known to be false, for example when we make such simplifying assumptions as ignoring gravitational interactions between planets when computing their orbits. I would argue that something similar can be said about the empirical disconfirmation of mathematical theories.

Let us look again at our examples. Although the success of the theory of complex numbers in physical applications is part of the story of their acceptance, it is not the whole story. Far more important was the role of the new numbers in other areas of mathematics; the problem solving power of complex numbers in analysis, their relation to the logarithmic and trigonometric functions and so on.[56]

In the same way, although the success of the calculus in physics was seen as providing some support for the new theory, it was the immense power of the theory to solve purely mathematical problems which carried the most weight. The calculus was initially adopted on the grounds that it provided a unified set of techniques for generating general solutions to mathematical problems which earlier mathematicians had only been able to solve for certain special cases – problems such as the construction of tangents to curves, calculation of arc-lengths and finding the maximum and minimum values.[57]

As the calculus developed, its applications in mathematics multiplied endlessly. As we saw in the previous chapter, Leibniz and Euler showed how the calculus provided a powerful set of techniques for finding the sums of infinite series.[58] Euler also showed how the calculus could cast some light on problems in number theory (the representation of

[55] See [Putnam 1974].
[56] See [Kline 1972, pp. 408-11,628-32, Nagel 1979].
[57] See [Kline 1972, pp. 34-77, Kitcher 1984, pp. 229-31].
[58] See [Kline 1972, pp. 436-67, Kitcher 1984, pp. 241-244].

integers as a sum of powers for example) and further extended the problem solving power of the calculus in geometry by showing how it could be applied to the study of surfaces. Mathematicians such as Gauss, Riemann and Legendre developed these ideas further, to create the new fields of differential geometry and analytic number theory.[59] Analysis is now such an indispensable part of mathematics, that mathematicians would not abandon it even if it turned out to be dispensable in every physical theory we now accept.

Although the utility of mathematical theories in science does count for something, it is utility in *mathematics* which counts even more. We saw some examples of this kind of utility in our discussion of Kitcher's evolutionary account of mathematics; the power of a new theory to solve important mathematical problems, or to give a more systematic, general or rigorous treatment of previous mathematical results.

Empirical evidence can be relevant to mathematics, but we do not reject whole mathematical theories on empirical grounds because there is also a great deal of *independent* confirming evidence for those theories; evidence derived from the utility of such theories in mathematics itself. In the face of an empirical disconfirmation of a scientific theory, we revise the auxiliary assumptions (such as 'physical space is Euclidean') rather than the mathematics. What I am suggesting is that we do so not because of a pragmatic desire to revise as little as possible of our total system of science, but because the mathematical theories are better supported by evidence than such auxiliary hypothesis.[60]

[59] See [Kline 1972, pp. 544-570, 829-33, 882-903].
[60] Of course, such auxiliary hypotheses are often not explicitly stated. This is especially obvious in the case of geometry, where the discovery and investigation of the non-Euclidean geometries led to a reinterpretation of geometrical theories which revealed for the first time, a distinction betwe n Euclidean geometry, considered as the true theory of a certain kind of abstract space and the auxiliary hypothesis that physical space is Euclidean. In a similar way, the apparent discrepancies in the positions of the moons of Jupiter led Roemer to

What the asymmetry between the empirical confirmation and disconfirmation of mathematical theories suggests is that empirical evidence is not the *only* kind of evidence we can have for a mathematical theory. The reason mathematical theories are not simply abandoned in the light of disconfirming *empirical* evidence is that there is also a great deal of *mathematical* evidence for those theories. In the next chapter, I will be taking a closer look at some of the varieties of this kind of mathematical evidence.

formulate and reject the previously implicitly assumed auxiliary hypothesis that the speed of light was instantaneous.

CHAPTER FIVE

MATHEMATICAL EVIDENCE

1. THEOREMS AND PROOFS

Consider the following example of a theorem from number theory; *every composite number is divisible by some prime number*. How do we know this? An obvious answer immediately suggests itself; we can prove it. Here is Euclid's proof from Book VII of the *Elements*:

> Let A be a composite number. I say that A is divisible by some prime number.
> For, since A is composite, some number will divide it. Let a number divide it and let it be B.
> Now if B is prime, what was enjoined will have been done.
> But if it is composite, some number will divide it. Let a number divide it, and let it be C.
> Then since C divides B and B divides A, therefore C also divides A.
> And if C is prime, what was enjoined will have been done. But if it is composite, some number will divide it.
> Thus, if the investigation be continued in this way, some prime number will be found which will divide the number before it, which will also divide A.
> For, if it is not found, an infinite series of numbers will divide the number A, each of which is less than the other: which is impossible in numbers.
> Therefore some prime number will be found which will divide the one before it, which will also divide A.
> Therefore any composite number is divisible by some prime number.[1]
>
> [Heath 1956, vol. II, p. 332]

In Book VII of the *Elements* Euclid does not, as in his presentation of geometry, state any axioms. He gives us only definitions; of prime number, composite number and so

[1] I have tinkered with the translation here, by substituting 'A is divisible by B' for 'A is measured by B'.

on.[2] In his proofs, of course, he does appeal to various principles which are neither proved nor even explicitly stated. Here for example, he appeals to the transitivity of divisibility and, in effect, to the principle that every set of natural numbers has a least element. This latter principle is equivalent to the principle of mathematical induction.

Euclid's argument is quite representative of the kind of proof we find in nearly all branches of mathematics up until the nineteenth century, especially in new and developing branches of mathematics. Cauchy's proof of Euler's conjecture on polyhedra is another good example. As we saw in chapter three, Cauchy's argument implicitly assumed at least three unproved lemmas concerning plane networks. Proofs of this kind do not begin from a set of clearly stated first principles or axioms. They may appeal to other theorems, but they may also implicitly depend on principles which are neither proved nor explicitly stated. The implicitly assumed premises of such a proof may even turn out to be false, as Lakatos points out. Clearly, proofs of this kind certainly do not guarantee the certainty of the theorem proved.

As a mathematical field develops and matures however, proofs become more rigorous. The principles on which a proof depends are stated clearly and if possible, proved. In this way, the first principles of the field (axioms and definitions) are gradually revealed and stated. A modern proof of Euclid's theorem for example, would appeal explicitly to the principle of mathematical induction and to other established theorems and axioms of number theory. A modern proof of Euler's conjecture (considered as a theorem concerning plane networks, rather than polyhedra) would appeal to established theorems and first principles of graph theory.

[2] See [Heath 1956, vol. II, pp. 277-8].

At least in a mature mathematical theory then, theorems are justified by deducing them from axioms. Of course, this immediately raises the question of how the axioms are justified. The descriptive approach to the epistemology of mathematics urges us to try to answer this question by examining the history and current practice of mathematics, in order to discover what kinds of evidence mathematicians actually cite when justifying their axioms. As we shall see, what such a study reveals is that axioms are very often justified by showing that they can be used to derive certain theorems.

2. Axioms and Definitions

2.1 Geometry

Consider Euclid's axiomatization of geometry. In Book I of the *Elements*, Euclid lays down five *postulates* and five *common notions*. The common notions (such as 'things equal to a third are equal to each other') are common to all sciences, while the postulates apply only to geometry. Euclid's postulates are as follows:[3]

(1) It is possible to draw a straight line from any point to any point.

(2) It is possible to extend a finite straight line continuously in a straight line.

(3) It is possible to draw a circle with any centre and radius.

(4) All right angles are equal to each other.

(5) If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, then the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

[3] See [Heath 1956, Kline 1972, p. 59].

From this apparently slender basis, Euclid was able to give proofs of an impressive number of theorems; on the construction of lines, angles and figures with ruler and compass, the properties of circles and triangles (including Pythagoras's theorem), as well as theorems on parallels, area, similar figures and solid geometry (including the classification of the five regular polyhedra). However, many historians of mathematics have argued that Euclid did not prove any results that were not already accepted by earlier geometers.[4] The great achievement of his work was not that he was able to prove anything new, rather it was the tremendous systematization his axioms bought to the subject. A motley collection of unrigorously demonstrated, unsystematically related results were shown to be derivable from a few simple principles. Euclid's axioms were justified by showing that they entailed, in a rigorous and systematic way, a large body of previously accepted results.

This is not to deny that Euclid's axioms have at various times been accorded the status of certain, self-evident truths. It is doubtful that this was Euclid's view of them however, or the view of his contemporaries. Recall that Euclid's geometry was proposed as a *physical* theory. In accordance with the usage of the word at the time, the *postulates* of the theory were intended as hypotheses to be tested by whether the results deduced from them agreed with physical reality.[5] The idea that the postulates were not simply hypotheses to be justified in terms of their consequences, but immutable self-evident truths only became widely accepted much later. No doubt this was due in part to the immense success of Euclidean geometry (considered as a scientific theory of the structure of physical space)

[4] See for example [Heath 1956, Kline 1972, pp. 56-7, Crowe 1988, p. 265].
[5] See [Kline 1972, p. 59, Lakatos 1976, p. 49].

but perhaps more significant was the absence of any alternative theories, which made it seem unthinkable that Euclid's postulates could be false.

This attitude was severely shaken by the discovery of the non-Euclidean geometries of course. I argued in the last chapter that this discovery led to a radical reinterpretation and reassessment of the status of geometry. From the new perspective, the axioms lost their status as self-evident truths and came to be seen as merely assumptions designed to yield the theorems of the geometry of a particular kind of abstract space. Since that time, self-evidence, or the lack of it, has gradually ceased to play any significant role in the justification of axioms.

Although Euclid's presentation of geometry was held up as a paradigm of rigour for many centuries, there are in fact many places where his proofs depend implicitly on assumptions (especially concerning the ordering of points on a line) which Euclid did not clearly state or prove.[6] These defects in Euclid's proofs, although sometimes noted, were not considered important until the discovery of the non-Euclidean geometries. Mathematicians then became acutely aware of the importance of clearly stating every assumption, no matter how obvious on which a proof depends, for such 'obvious' assumptions may fail to hold in the non-Euclidean geometries. This led eventually to a more rigorous axiomatic reconstruction of the Euclidean and non-Euclidean geometries.

This was the motivation for Hilbert's axiomatization of geometry, mentioned in the previous chapter.[7] By giving an entirelv formal axiomatization of the relations of betweenness of points and congruence of lines, Hilbert was able to provide entirely rigorous proofs of Euclidean theorems. He was also able to show that simply by replacing

[6] See [Kline 1972, pp. 1005-7] for some examples.
[7] See [Hilbert 1899]

Euclid's parallel axiom with the Bolyai-Lobachevsky axiom, we obtain the axioms of hyperbolic geometry. To obtain the axioms of spherical geometry (which had not previously been treated axiomatically at all) we replace the parallel axiom with an axiom stating that there are *no* parallels to a given line through a point not on that line. In fact there are two possible such geometries; *single* elliptic geometry (where any two lines meet at exactly one point) and *double* elliptic (where any two lines meet at two points). The straight lines of these geometries have the properties of circles, so we must also modify the axioms governing the relations of order between points on a line (the axioms of betweenness), so that they describe the order relations of points on a circle.

Again, Hilbert did not prove anything new about these geometries. His axioms were justified by showing that they entailed, in a rigorous way, all the accepted theorems of the Euclidean and non-Euclidean geometries.[8]

2.2 Analysis

The high standard set by Euclid in the *Elements*, where theorems are justified by deducing them from explicitly stated axioms, was more or less ignored by mathematicians until the nineteenth century. We have seen how the discovery of non-Euclidean geometries at that time led mathematicians to give a more rigorous axiomatic reconstruction of geometry. Another nineteenth century development was of course, the project of constructing a rigorous foundation for analysis.

I described the beginnings of that story in chapter three. Recall that one motivation for Cauchy's definitions of continuity and convergence was to establish a certain result – that the sum of a convergent series of continuous functions is continuous – as a means of

[8] See [Kline 1972, pp. 1005-7].

resolving the Fourier question. Further support for Cauchy's reconstruction of analysis in terms of the limit concept came from his definition of the derivative of a function, which allowed him to give a rigorous and systematic derivation of the standard rules for differentiation. In this case, the first principles take the form of definitions rather than axioms, but they too are justified in terms of their consequences, by showing how they can be used to deduce already accepted results or a previously unproved conjecture.

Of course, as we saw in chapter three, Cauchy's proof of the conjecture on convergent series of functions was known to be faulty, since the 'theorem' admitted counter-examples. Cauchy's definitions also proved to be inadequate in another way. Cauchy had stated a certain condition for a series to converge to a limit, but was unable to prove that the condition was sufficient for convergence. The problem is that Cauchy's definitions cannot be used to establish the *existence* of limits of series. When he needed to establish such a limit, Cauchy would often fall back on geometrical intuition. For example, in the *Cours d'Analyse* he gives the 'geometrical' argument for the intermediate value theorem mentioned in chapter one. Although Cauchy attempted a 'purely analytic' demonstration of the theorem, his proof broke down at a crucial point, where he implicitly assumed the condition for convergence which he had stated but been unable to prove.[9]

What these problems with Cauchy's reconstruction of analysis revealed was the necessity of giving an adequate analysis of the real numbers. What was needed was a characterisation of the real number system which would entail the existence of real numbers satisfying certain conditions, such as Cauchy's condition for convergence and in this way establish the required limit existence theorems.

[9] For further details see [Kline 1972, p. 963, Kitcher 1984, pp. 259-263].

Dedekind set about this task in 'Continuity and the Irrational Numbers' [Dedekind 1872].[10] He begins by asking what is meant by the *continuity* of the points on a straight line. It cannot just be that between any two points there is a third, for the set of all rational points on the line also has this property but is not continuous since it contains 'gaps' corresponding to irrational numbers such as $\sqrt{2}$. Dedekind argued that the continuity of the points on a straight line depends on the following principle: '[i]f all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions' [ibid. p. 11].

Dedekind proposed to extend this definition of geometrical continuity to characterise the real numbers. He introduces the idea of a *cut* on the set of rational numbers *R*. A cut is a separation of the set of all rational numbers into two classes A_1 and A_2 such that any number in A_1 is less than every number in A_2. By analogy with his principle of continuity, Dedekind proposes that we can introduce the real numbers as the unique entity 'corresponding to' or 'produce by' each such cut $\langle A_1, A_2\rangle$. In fact, we can simply identify the real numbers with these pairs of sets of rationals. For some cuts, there is either a largest element in A_1 or a smallest element in A_2 and these cuts represent those real numbers which are also rational. For example the cut defined by $A_1 = \{x \in R: x < \frac{1}{2}\}$ represents the real number 0.5. Here A_2 (the set of all rationals not in A_1) contains a smallest element, the rational number ½. But not every cut corresponds to a rational number. The cut defined by $A_1 = \{x \in R: x^2 < 2\}$ for example, represents the irrational number $\sqrt{2}$.

[10] Cantor gave an equivalent account of the real number system using a different construction to Dedekind's. See [Kline 1972, pp. 982-5] for details.

Having introduced the real numbers as cuts on the rationals, Dedekind shows how to define an ordering relation on the reals, in terms of a corresponding relation on cuts. Using this definition Dedekind is able to prove the *law of trichotomy* for real numbers:

For any two real numbers a and b, either $a < b$, $a = b$ or $b < a$

By defining the operations of addition and multiplication on cuts, Dedekind is able to prove such familiar properties as the associative and commutative laws for the addition and multiplication of real numbers. He then proceeds to establish the continuity of the real numbers, by proving that any division of the real numbers into two classes X_1 and X_2 such that every member of X_1 is less than every member of X_2 corresponds to a unique real number.

Dedekind concludes his book by showing how his characterisation of the real numbers can indeed be used to prove those theorems on the existence of limits which Cauchy had failed to establish. In particular, he is able to give an elegant proof of the theorem that any sequence of real numbers which 'grows continuously but not beyond all limits' approaches a limiting value [ibid. pp. 25-7]. That is if r_1, r_2, r_3 .... is a sequence of real numbers such that for all n, $r_n < r_{n+1}$ and if there is a real number a such that for all n, $r_n < a$ then there is a real number b such that $\lim_{N\to\infty}\sum_{n=1}^{N} r_n = b$. This theorem fills the gap in Cauchy's proof of the intermediate value theorem.[11]

Once again we see a set of first principles justified by their consequences. Dedekind justifies his definitions by showing how they can be used to establish previously accepted, but unproved properties of the real numbers; the continuity of the real number system, the

[11] See also [Kline 1972, pp. 985-6, Kitcher 1984, pp. 264-8].

standard rules for addition and multiplication of real numbers and theorems on the existence of limits of sequences of real numbers.

2.3 NUMBER THEORY

Dedekind's construction assumes the rational numbers as given. Since the rational numbers are ratios of natural numbers, the obvious next step was to show how the natural numbers themselves could be characterised. As we saw in chapter one, Dedekind gave an account of the natural numbers in *Was sind und was sollen die Zahlen?* [Dedekind 1888]. Building on Dedekind's work, Giuseppe Peano gave the first axiomatic treatment of the natural numbers in his book *Arithmetices principia, nova methodo exposita* ('The Principles of Arithmetic, Presented by a New Method') [Peano 1889]. Peano begins with the undefined concepts 'set', 'belongs to', '1', 'natural number' and 'successor'. He then propose the following axioms governing these last three concepts:

(1) 1 is a natural number.

(2) 1 is not the successor of any natural number.

(3) Every natural number has a successor.

(4) If the successor of a is equal to successor of b, then $a = b$.

(5) If a set A of natural numbers contains 1, and if whenever A contains any natural number a, it also contains the successor of a, then A contains all the natural numbers.

Compare this list to the formulation of the Dedekind-Peano axioms given in chapter one. Notice that Peano starts with the number one, rather than with zero and that axiom (5), the principle of mathematical induction, is stated in terms of *sets* of numbers. Peano then gives the now standard recursive definitions of the operations of addition and multiplication on the natural numbers:

(Def.+) $a + 1 = \mathrm{S}(a)$

$a + \mathrm{S}(b) = \mathrm{S}(a + b)$

(Def. ×) $a \times 1 = a$

$a \times \mathrm{S}(b) = (a \times b) + a$

where $\mathrm{S}(x)$ is the successor of x. From these definitions and the above axioms, Peano went on to prove various familiar results such as:

(1) $a + b = b + a$
(2) $a \times b = b \times a$
(3) $a + (b + c) = (a + b) + c$
(4) $a \times (b \times c) = (a \times b) \times c$
(5) $a \times (b + c) = (a \times b) + (a \times c)$

Given the natural numbers and their properties, the positive and negative integers can be defined as ordered pairs of natural numbers; if a and b are natural numbers, then the ordered pair $\langle a, b\rangle$ represents the integer $a - b$. With appropriate definitions of addition and subtraction, we obtain all the usual properties of the integers. The rational numbers can then be defined as ordered-pairs of integers; if a and b are integers, then the ordered pair $\langle a, b\rangle$ represents the rational number $\frac{a}{b}$. Again, with appropriate definitions of the operations of addition and multiplication of such pairs, we obtain all the usual properties of the rationals. Given the rationals, the real numbers can then be defined as pairs of sets of rational numbers along the lines suggested by Dedekind. In this way the nineteenth century mathematicians hoped to build up to a rigorous treatment of the real numbers which would enable them to give rigorous proofs of the fundamental theorems of analysis.

The nineteenth century mathematicians showed how their characterisations of the real numbers, rationals, integers and natural numbers could be used to derive the basic properties of those number systems; properties which were of course, already accepted and never in any doubt. It was the power to prove certain already accepted results that gave these axioms and definitions their justification. Morris Kline makes the point like this:

> The rigorization of mathematics may have filled a nineteenth century need, but it also teaches us something about the development of the subject. The newly founded logical structure presumably guaranteed the soundness of mathematics; but the guarantee was somewhat of a sham. Not a theorem of arithmetic, algebra or Euclidean geometry was changed as a consequence, and the theorems of analysis had only to be more carefully formulated. In fact, all that the new axiomatic structures and rigor did was substantiate what mathematicians knew had to be the case. Indeed the axioms had to yield the existing theorems rather than determine them.
>
> [Kline 1972, p. 1026]

2.4 SET THEORY

Set theory was invented by Georg Cantor in a series of papers spanning a period of time from 1874 to 1897.[12] Some of Cantor's earliest mathematical work was in a field which developed out of research on the Fourier question; the representation of functions by trigonometric series. In particular, Cantor wanted to state conditions for the *uniqueness* of such a representation. He first proved that a function is uniquely represented by a trigonometric series if the series converges at every point. He then generalised this result to series that converge at all but a finite number of points and then to series that fail to converge at an infinite number of exceptional points. These sets of exceptional points form

[12] Cantor's final and quite accessible summary of his theory [Cantor 1895-7] has been translated into English by P. E. B Jourdain [Cantor 1955].

a sequence; infinite sets with one accumulation point, infinite sets with finitely many accumulation points and so on.[13]

Cantor's work on these infinite sequences of sets of real numbers led him to an investigation of the properties of infinite sets in general and ultimately to his theory of transfinite cardinal and ordinal numbers.[14] Now the notion of an infinite set had puzzled mathematicians for centuries. Galileo, for example in his *Dialogues Concerning Two New Sciences* noted that the points on two lines can be put in one-one correspondence, which suggests there are the same number of points on each, even when one line is longer than the other. He also pointed out that the set of positive integers can be put in one-one correspondence with their squares, even though the set of the squares of positive integers is only a proper part of the set of all positive integers.[15] These sorts of puzzles led many mathematicians to reject the notion of a completed infinite totality; only the *potential* infinite was to be allowed. However, the nineteenth century work on the rigorization of analysis had revealed the need to assume the existence of various infinite sets; infinite sets of rational numbers in Dedekind's construction of the real numbers, for example.

Cantor's key insight was to use the notion of a one-one correspondence, which Galileo and others had rejected, as the condition for two sets to be of the same size. He used this idea to define the notion of the *power* of a set; a set A has the same power as a set B ($A \approx B$) if and only if there is a one-one correspondence between the elements of A and the elements of B. On the other hand, the power of A is greater than the power of B ($A \succ$

[13] A point p of a set S is an accumulation point iff every interval containing p contains infinitely many points in S.

[14] For further details of Cantor's early work on trigonometric functions and its relation to his transfinite set theory, see [Jourdain 1955, pp. 24-37, Maddy 1997, pp. 15-17].

[15] See [Galileo 1952, pp. 18-40, Kline 1972, p. 993].

B) if and only if B can be put in one-one correspondence with a subset of A, but A cannot be put into one-one correspondence with a subset of B. That is:

(Def $\succ$): $A \succ B \leftrightarrow \exists C(C \subseteq A \,\&\, B \approx C) \,\&\, \sim\exists C(C \subseteq B \,\&\, A \approx C)$

This was immensely significant, because Cantor was then able to show that not all infinite sets have the same power. In particular, he was able to show that the power of the set of real numbers is greater than the power of the set of natural numbers[16]. On the other hand, he proved that there are just as many rational numbers as natural numbers, by showing how to match each rational number to a unique natural number. Even more surprisingly he was able to show that the set of real numbers has the same power as the set of all *n*-tuples of real numbers and hence that the set of points on a line is the same size as the set of points in any *n*-dimensional space.[17]

The concept of the power of a set suggested to Cantor a generalization of the arithmetic of the natural numbers, which would allow for infinite or *transfinite* numbers. He introduces this theory with the following words:

> The description of my investigations in the theory of aggregates has reached a stage where their continuation has become dependent on a generalization of the real positive integers beyond the present limits; a generalization which takes a direction in which, as far as I know, nobody has yet looked.
>
> I depend on this generalization of the number concept to such an extent that without it I could not freely take even small steps forward in the theory of sets. I hope that this situation justifies, or if necessary excuses the introduction of seemingly strange ideas into my arguments. In fact the purpose is to generalize or extend the series of real integers beyond infinity. Daring as this might appear, I express not only the hope but also the firm conviction that in due course this generalization will be acknowledged as a quite simple, natural step.
>
> [Cantor 1883, cited in Kline 1972, p. 998]

[16] Cantor gave two proofs of this theorem. The one most commonly used today is the famous 'diagonal argument'. See [Kline 1972, p. 997, Stewart 1987, p. 59, Penrose 1989, p. 84] for details.

[17] Cantor apparently spent three years trying to prove that such a correspondence was impossible. On finally proving that there is such a correspondence, he wrote in a letter to Dedekind, 'I see it but I do not believe it'. [cited in Kline 1972, p. 997].

Cantor proposes then to think of the power of a set as a *number*. Hence the cardinal number of a set A = the cardinal number of a set B if and only if there is a one-one correspondence between the elements of A and the elements of B.[18] If we introduce the symbol CARD(A) for the cardinal number of a set A then we have:

$$\mathrm{CARD}(A) = \mathrm{CARD}(B) \leftrightarrow A \approx B$$

The relation of one set being greater in power than another can then be used to define an ordering relation on these numbers.

$$\mathrm{CARD}(A) > \mathrm{CARD}(B) \leftrightarrow A \succ B$$

Cantor was also able to show that for any cardinal, there is a greater one. Given a set A, we can consider the *power set* of A, *P*(A), which is the set of all subsets of A. Cantor proved that the elements of *P*(A) always outnumber the elements of A, by showing that it is impossible for there to be a one-one correspondence between the elements of A and the elements of *P*(A).[19] But since one subset of *P*(A) consists of the unit set of each element of A, it is obvious that A can be placed in one-one correspondence with a subset of *P*(A). Hence:

(1) $\mathrm{CARD}(P(A)) > \mathrm{CARD}(A)$

[18] See [Cantor 1955, p. 87]. Notice the formal similarity between Cantor's notion of the cardinal number of a set and Frege's notion of the number belonging to a concept; the number of a concept F is the same as the number of a concept G if and only if there is a one-one correspondence between the Fs and the Gs. Cantor in fact attempted a proof of his principle, by defining the cardinal number of a set to be 'the general concept which, by means of our active faculty of thought, arises from the [set] when we make abstraction of the nature of its various elements...and of the order in which they are given' [ibid. p. 86]. If we start with another set containing equally many members as the first, then we will arrive at the same abstract set by means of this operation of abstraction. This conception of an 'active faculty of the mind' which can create objects in this way was shared by many philosophers and mathematicians of the time, including Husserl and Dedekind. The idea was vigorously demolished by Frege in the *Grundlagen* [1884, §§29-44]. See also [Dummett 1991, pp. 82-7,50-2].

[19] For the details of Cantor's proof see [Kline 1972, p. 999, Suppes 1972, p. 97]. A clear informal version of the argument can be found in [Quine 1987, p. 96].

Cantor also showed how to define analogs of the arithmetical operations on these new numbers[20]. The sum of two cardinal numbers a and b is the cardinal number of the union of any two disjoint sets whose cardinal numbers are a and b. The product of a and b is the cardinal number of the cartesian product of any two sets whose cardinal numbers are a and b. Finally a^b is the cardinal number of the set of all functions from a set B to a set A, where $\mathrm{CARD}(A) = a$ and $\mathrm{CARD}(B) = b$.

The cardinal number of a *finite* set is a finite or natural number. If we restrict our attention to finite sets, then Cantor's cardinal arithmetic corresponds exactly to the ordinary arithmetic of the natural numbers. In fact, if we define the number zero and the successor operation on cardinal numbers as follows:

$0 = \mathrm{CARD}(\varnothing)$

$S(a) = \mathrm{CARD}(A')$ - where $a = \mathrm{CARD}(A)$ and $A' = A \cup \{A\}$

then we can very easily prove all the Dedekind-Peano axioms for number theory.[21] Cantor remarks that in this way he has given 'the most natural, shortest, and most rigorous foundation for the theory of finite numbers' [Cantor 1955, p. 98].

The cardinal number of an *infinite* set is a transfinite number. Cantor introduces the symbol $\aleph_0$ (aleph-null) for the cardinal number of the set of all natural numbers and the symbol c for the cardinal number of the set of all real numbers – the cardinality of the

[20] See [Cantor 1955, pp. 91-97].
[21] See [Cantor 1955, pp. 97-103, Suppes 1972, pp. 121-3]. The fact that the Dedekind-Peano axioms can be proved from the principle $\mathrm{CARD}(A)=\mathrm{CARD}(B) \leftrightarrow A \approx B$ is perhaps not too surprising given the analogy between this principle and Hume's Principle, which also implies those axioms, as Frege demonstrated in the *Grundlagen*.

continuum[22]. Since the set of all real numbers has a greater power than the set of all natural numbers, we have:

(2) $c > \aleph_0$

On the other hand, we can show that the set of all real numbers can be put in one-one correspondence with set of all functions from natural numbers into the set $\{0, 1\}$.[23] By the definition of the operation of exponentiation on cardinal numbers, this means that:

(3) $c = 2^{\aleph_0}$

The arithmetic of transfinite numbers is very different to that of the natural numbers. Although many of the properties of addition and multiplication (such as the commutative and associative properties) continue to hold, others apply only to the case of finite cardinal numbers. For example, the principle that for all non-zero a and b, $a + b > a$ applies only when a and b are finite numbers. If a and b are transfinite numbers, although we always have $a + b \geq a$, it may happen that $a + b = a$. In particular, Cantor shows that we have $\aleph_0 + \aleph_0 = \aleph_0$. Given this, if we multiply both sides of (3) by c we have:

$$c \cdot c = 2^{\aleph_0} \cdot c = 2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0 + \aleph_0} = 2^{\aleph_0} = c$$

Hence, by repeated multiplication by c, we obtain:

$$c^n = c$$

where n is any finite cardinal number. This means that the set of points in any n-dimensional space has the same power as the linear continuum, a result which, as mentioned above, Cantor had already obtained in his early work on powers. In the same

[22] See [Cantor 1955, pp. 96,104].
[23] To see this notice that any function from the set of natural numbers into the set {0,1} specifies an ordered sequence of ones and zeros. Every real number, when written out in binary notation, corresponds to such a sequence and conversely every such sequence specifies a unique real number. See [Cantor 1955, p. 96].

way, Cantor easily establishes that $\aleph_0 \cdot \aleph_0 = \aleph_0$ and then, raising both sides of (3) to the power of $\aleph_0$, he obtains:

$$c^{\aleph_0} = \left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0} = c$$

which means that the set of points in any *infinite* dimensional space also the same power as the linear continuum. Cantor remarks, '[t]hus the whole contents of my paper in Crelle's *Journal* ...are derived purely algebraically with these few strokes of the pen from the fundamental formulae of the calculation with cardinal numbers'. [ibid. p. 97].

Cantor then introduces the *ordinal* numbers. An *order structure* (or ordered set) is a pair $\langle A,R \rangle$ where A is set and R is an ordering relation on A[24]. Two order structures $\langle A,R \rangle$ and $\langle B,S \rangle$ are *similar* ($\langle A,R \rangle \equiv \langle B,S \rangle$) if and only if the elements of A can be put into one-one correspondence with the elements of B in such a way that the order relations are preserved. That is, there is a 1-1 function $f(x)$ from A to B such that aRb if and only if $f(a)Sf(b)$. Two order structures which are similar are said to have the same *order type*. Cantor then introduces the ordinal numbers as the order types of well-ordered sets, by means of the principle:

$$\text{ORD}(\langle A,R \rangle) = \text{ORD}(\langle B,S \rangle) \leftrightarrow \langle A,R \rangle \equiv \langle B,S \rangle$$

That is, the ordinal number of an well-ordered set $\langle A,R \rangle$ is equal to the ordinal number of a well-ordered set $\langle B,S \rangle$ if and only if $\langle A,R \rangle$ is similar to $\langle B,S \rangle$.[25]

[24] R must be a *simple ordering* on A. That is: (i) any two elements of A have a definite order, so that for all $a,b \in A$ either aRb or bRa. (ii) R is transitive on A – for all $a,b,c \in A$ if aRb and bRc then aRc. For the introduction of the ordinal numbers, Cantor also requires that R is a *well-ordering* of A, which means that every non-empty subset of A contains a least element in the ordering provided by R.
[25] See [Cantor 1955, pp. 112-13, 151-2].

If we take finite sets of natural numbers in their usual ordering, we obtain the *finite ordinals*:

$$1 = \text{ORD}(\langle \{0\}, < \rangle)$$

$$2 = \text{ORD}(\langle \{0,1\}, < \rangle)$$

$$3 = \text{ORD}(\langle \{0,1,2\}, < \rangle)$$

and so on. Cantor introduces the symbol ω for the ordinal number of the set N of all finite cardinal numbers, taken in their usual order.[26] That is:

$$\omega = \text{ORD}(\langle N, < \rangle)$$

Cantor then shows how we can define an ordering relation on the ordinal numbers themselves. If a is any element of a well-ordered set $\langle A,R \rangle$ then the *segment* of A determined by a is the set of all elements in A which are less than a, according to the ordering defined by R. We can then define a relation < on the ordinal numbers in an analogous way to the corresponding definition for cardinal numbers. If $\alpha = \text{ORD}(\langle A,R \rangle)$ and $\beta = (\langle B,S \rangle)$ we have: [27]

$$\alpha < \beta \text{ iff there is a segment C of B such that C is similar to A}$$

Hence for the finite ordinals 1, 2, 3,... we have $1 < 2 < 3$ ..... since there is a segment of $\langle \{0,1\}, < \rangle$ which is similar to $\langle \{0\}, < \rangle$ and a segment of $\langle \{0,1,2\}, < \rangle$ which is similar to $\langle \{0,1\}, < \rangle$. Since there is a segment of ω which is similar to any finite ordinal n (namely the segment determined by n itself) then we have $n < \omega$ for every finite ordinal n.

Cantor also defines operations of addition and multiplication on ordinal numbers, in terms of corresponding operations on ordered sets.[28] The resulting *ordinal arithmetic* is

[26] See [Cantor 1955, p. 115]
[27] See [ibid. pp. 141, '52-3]
[28] See [ibid. pp. 119-122, 153-6]

even less like the arithmetic of the natural numbers than cardinal arithmetic[29]. In particular, ordinal addition and multiplication are not commutative. For example, $1 + \omega \neq \omega + 1$, since $1 + \omega = \omega$, but $\omega + 1 > \omega$. Likewise $2 \times \omega \neq \omega \times 2$, since $2 \times \omega = \omega$, but $\omega \times 2 = \omega + \omega$ and $\omega + \omega > \omega$[30].

Cantor classifies ordinal numbers according to the size of the corresponding ordered sets and this provides him with a useful means of defining the higher transfinite cardinal numbers. Consider first the set of all finite ordinals, Z_0:

$$Z_0 =_{df} \{\alpha\text{: } \alpha = \text{ORD}(\langle A,R\rangle) \text{ and A is finite }\}$$

We can then define aleph-null to be the cardinal number of this set:

$$\aleph_0 =_{df} \text{CARD}(Z_0)$$

Now consider the set Z_1 of all countable ordinals:

$$Z_1 =_{df} \{\alpha\text{: } \alpha = \text{ORD}(\langle A,R\rangle) \text{ and } A \approx Z_0\}$$

Equivalently:

$$Z_1 =_{df} \{\alpha\text{: } \alpha = \text{ORD}(\langle A,R\rangle) \text{ and } \text{CARD}(A) = \aleph_0\}$$

Z_1 is the set of all ordinal numbers of ordered sets which can be put in one-one correspondence with the set of all finite ordinals[31]. Z_1 includes the ordinals ω, $\omega+1$, $\omega+2$ and so on, since each of these is the ordinal number of a set whose cardinal number is $\aleph_0$.

We can then define:

[29] Although of course the arithmetic of the *finite* ordinals is exactly the same as the standard arithmetic of the natural numbers.

[30] More generally, for any finite ordinal $n \geq 1$, we have $n + \omega \neq \omega + n$, since $n + \omega = \omega$ for any finite ordinal n, but $\omega + n > \omega + (n\text{-}1)$. Likewise, for any finite ordinal $n \geq 1$, we have $n \times \omega \neq \omega \times n$, since $n \times \omega = \omega$, but $\omega \times n > \omega \times (n\text{-}1)$. See [Cantor 1955, p. 163-4]. See also [Suppes 1972, pp. 195-238] for a modern treatment of transfinite ordinal arithmetic.

[31] Cantor calls Z_0 the 'first number class' and Z_1 the 'second number class' [Cantor 1955, pp. 159-60].

$$\aleph_1 =_{df} \text{CARD}(Z_1)$$

That is, $\aleph_1$ is the cardinal number of the set of all countable ordinals. In general, we can define:

$$Z_{n+1} =_{df} \{\alpha\text{: } \alpha = \text{ORD}(\langle A,R\rangle) \text{ and } A \approx Z_n\}$$

$$\aleph_n =_{df} \text{CARD}(Z_n)$$

Cantor was able to prove[32] that $\aleph_1$ is not only larger than $\aleph_0$ but that it is also the *next* largest cardinal number after $\aleph_0$, in the sense that there is no transfinite cardinal greater than $\aleph_0$ but less than $\aleph_1$. That is:

(4) $\quad \aleph_0 < \aleph_1$ and $\sim\exists a(\aleph_0 < a < \aleph_1)$

We know that Z_0 – the set of all finite ordinals – is the same size as the set of all natural numbers (since $Z_0 \approx N$) and therefore, $\aleph_0 = \text{CARD}(N)$. How big is Z_1? Cantor conjectured that in fact Z_1 is the same size as the set of all real numbers, that is:

(CH) $\quad \aleph_1 = c$

This conjecture, that the cardinality of the set of all countable ordinals is equal to the cardinality of the continuum has become known as the *continuum hypothesis*. Given (2) and (4) above, we can show that (CH) is equivalent to the claim that there is no transfinite cardinal greater than $\aleph_0$ but less than c:

[32] See [Cantor 1955, pp. 169-73]

(CH') $\sim\exists a(\aleph_0 < a < c)$

In terms of sets, (CH') says that there is no infinite set which is bigger than the set of all natural numbers, but smaller than the set of all real numbers. This in turn implies that there are exactly two kinds of infinite sets of real numbers; those which contain countably many real numbers and those which can be put in one-one correspondence with the set of all real numbers.[33] Given (3) above, it is obvious that (CH) is also equivalent to the claim:

(CH'') $\quad \aleph_1 = 2^{\aleph_0}$

which is the usual modern formulation of the continuum hypothesis.[34] Despite many efforts, Cantor was unable to prove the conjecture and it remains unresolved to this day.

Cantor's work was highly original and revolutionary. In fact, it was so revolutionary that many mathematicians refused to accept it. Poincaré, for example, remarked that '[l]ater generations will regard [Cantor's set theory] as a disease from which one has recovered' and Herman Weyl described the hierarchy of alephs as 'a fog on a fog'.[35] The most vehement opponent of Cantor's set theory was Leopold Kronecker who attacked Cantor's ideas for many years and left many mathematicians suspicious of it. Nonetheless, as the tremendous power of set theory in numerous branches of mathematics became clear, the theory began to gain acceptance.

An early success for Cantor's theory was the explanation of the existence of transcendental numbers. A number is said to be *algebraic* if it satisfies some polynomial

[33] This last conjecture was stated by Cantor at the end of an early paper on infinite sets of the same power as the continuum, but he was unable to prove it [Cantor 1878, see also Jourdain 1955, p. 45]. He stated the continuum hypothesis in the form (CH) and noted that it implied the conjecture of the 1878 paper in a letter to Mittag-Leffler. [See Moore 1989, p. 85, Maddy 1997, pp. 63-4].

[34] The generalized continuum hypothesis is that for all α: $\aleph_{\alpha+1} = 2^{\aleph_\alpha}$ [see Gödel 1947, p. 473].

[35] Cited in [Kline 1972, p. 1003].

equation with rational coefficients. Thus $\sqrt{2}$ is an algebraic number since it satisfies the equation $x^2 - 2 = 0$. A number is *transcendental* if it is not algebraic, that is, if it is not the root of any polynomial equation with rational coefficients. Until 1844, it was not known whether there were any transcendental numbers at all. In that year however, Liouville proved that any number of the form $\frac{a_1}{10} + \frac{a_2}{10^{2!}} + \frac{a_3}{10^{3!}} + \frac{a_4}{10^{4!}} + \ldots$ where the a_i are arbitrary integers from 0 to 9 is transcendental. So in particular, the number:

$$\frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^6} + \frac{1}{10^{24}} + \ldots = 0.110001000000000000000001\ldots\ldots$$

is transcendental. Then in 1873, Charles Hermite gave a proof that e is transcendental. It had long been conjectured that π was also transcendental, but Hermite despaired of finding a proof. Ferdinand Lindemann gave the first proof that π is transcendental in 1882.[36] In this way mathematicians came to accept the existence of transcendental numbers although the proofs involved were complex and not highly explanatory.

Now Cantor was able to supply a simple and explanatory proof of the existence of transcendental numbers, without constructing a single specific example. He was able to prove that the set of algebraic numbers can be put in one-one correspondence with the set of natural numbers. But since there are more real numbers than natural numbers, it follows immediately that there are real numbers which are not algebraic and which are therefore

[36] Lindemann was able to establish that if x is algebraic then $e^x + 1 \neq 0$. But by Euler's famous result we have $e^{i\pi} + 1 = 0$. Hence the number $i\pi$ is not algebraic. But since i is algebraic (it satisfies the equation $x^2 = -1$) and the product of two algebraic numbers is algebraic, it follows that π is not algebraic. Hence π is transcendental. See [Kline 1972, pp. 981-2].

transcendental. In fact, there must be uncountably many transcendental numbers, since there are uncountably many more real numbers than natural or algebraic numbers.[37]

Thus Cantor was able to prove, in an elegant and explanatory manner, a known result. In this way, his theory received an indirect degree of justification. More and more of this sort of evidence began to accumulate to the theory of transfinite numbers; its important applications in analysis, algebra, topology and measure theory for example[38], until in 1926 Hilbert was able to say "no one shall drive us out of the paradise which Cantor has created for us" [Hilbert 1926, p. 191].

Cantor's set theory was not axiomatic. The first axioms for set theory were proposed by Ernst Zermelo.[39] Zermelo had at least two motives for axiomatizing set theory. One motivation was to avoid the paradoxes. Cantor himself had already noted that some sets can have inconsistent properties. Consider the set of all sets, U. Now consider the set of all subsets of U, *P*(U). By Cantor's theorem (1) above, we know that CARD(*P*(U)) > CARD(U). But clearly, every member of *P*(U) is a set. Therefore, every member of *P*(U) is also a member of U – since U is the set of *all* sets. This means that *P*(U) is a subset of U. But it is easy to show that if A is a subset of B, then CARD(A) ≤ CARD(B). Hence, CARD(*P*(U)) ≤ CARD(U). This is a contradiction: we have shown that *P*(U) is both larger than U and less than or equal to U.[40]

[37] For a discussion of Cantor's proof that there only countably many algebraic numbers see [Kline 1972, pp. 996-7 and Stewart 1987, pp. 59-60].
[38] Some of these applications are described in [Kline 1972, p. 1003;1040-52;1158-82].
[39] In [Zermelo 1908].
[40] Cantor also pointed out that the set of all cardinal numbers would have a cardinal number greater than any cardinal. But there cannot be such a cardinal number because for any cardinal number, there is a greater one. Likewise, Burali-Forti pointed out that the set of all ordinal numbers, since it can be well-ordered, would have an ordinal number, but this ordinal number would be greater than every ordinal, including itself. [Burali-Forti 1897].

The most famous of the paradoxes was formulated by Russell and is analogous to his derivation of the contradiction from Frege's Axiom V. In terms of sets (as opposed to extensions of concepts) Russell's paradox can be stated in the following way. Consider the set of all sets which are not members of themselves. Call this set R. By the definition of R, a set *a* is an element of R if and only if *a* is not an element of *a*. If we take *a* to be R itself, then we have: R is and element of R if and only if R is not an element of R. Contradiction.[41]

By this time the concept of a set had come to be seen as indispensable in mathematics. As we have seen, the nineteenth century mathematicians made frequent use of the concept in their reconstruction of analysis on the basis of arithmetic. Cantor's work suggested that this entire development could be carried out from within the framework of pure set theory. From within this framework, the basic concepts of mathematics, such as 'number' and 'function' could be precisely defined and the fundamental principles of the number systems and of analysis could all be given rigorous proofs. In this way, the idea of set theory as a foundation for the entire body of classical mathematics began to emerge.[42] Given this climate of opinion, the discovery of the paradoxes came to be seen as a very serious impediment to the progress of mathematics.

Before the paradoxes, mathematicians had simply assumed without proof that certain sets exist. In particular, they had made frequent use of definition by abstraction, whereby a set is defined as the collection of all objects having a certain property. Cantor made frequent use of such definitions in his development of set theory (for example: the

[41] See [Russell 1903, p. 101].
[42] Hence Cantor's remark that 'pure mathematics... is nothing other than pure set theory' [Cantor 1884, p. 84]. Many other mathematicians came to see set theory in a similar way. See for example [Zermelo 1908, p. 200, Hilbert 1926, p. 191].

power set of A = the set of all subsets of A) and the nineteenth century mathematicians had also used them in their reconstruction of analysis on the basis of arithmetic (for example: $\sqrt{2}$ = the set of all rational numbers whose squares are less than 2). No one had thought it was necessary to prove that a set defined by abstraction actually exists, but the paradoxes had shown that some definitions of this form can lead to contradiction (for example: A = the set of all sets). On the other hand, assumptions about the existence of certain sets had become indispensable in mathematics. The problem then was to find criteria governing the existence of sets which would allow mathematicians to avoid the 'inconsistent' sets and rigorously establish the existence of all the sets required for classical mathematics.

Zermelo's approach to this problem was to take the existent body of Cantorian set theory and find axioms which would entail everything that was useful and well founded in the theory but which would not generate any of the known paradoxes:

> ... in view of the "Russell antinomy" of the set of all sets that do not contain themselves as elements, it no longer seems admissible today to assign to an arbitrary logically definable notion a set, or class, as its extension....Under these circumstances there is at this point nothing left for us to do but to proceed in the opposite direction and, starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline. In solving the problem we must, on the one hand, restrict these principles sufficiently to exclude all contradictions and, on the other, take them sufficiently wide to retain all that is valuable in this theory.
>
> [Zermelo 1908, p. 200 in van Heijenoort 1967]

It is clear from what Zermelo says here that his axioms are to be justified in terms of what they enable us to prove. The axioms must be strong enough to enable us to prove everything valuable in set theory while at the same time they must not be so strong as to entail the paradoxes.

According to Zermelo, the principle that leads to the paradoxes is the principle of *naïve comprehension*. This principle states that for any condition there is a set which consists of every object which meets that condition:

(NC) $\exists x \forall y(y \in x \leftrightarrow \phi(y))$

Here the quantifiers are assumed to be restricted to sets. Given (NC) we can derive Russell's paradox as follows. Let $\phi(y)$ the formula $y \notin y$. Then by (NC) we have:

(1) $\exists x \forall y(y \in x \leftrightarrow y \notin y)$

This says that there is a set x with a certain property. Call that set R. Then we have:

(2) $\forall y(y \in \mathrm{R} \leftrightarrow y \notin \mathrm{y})$

Hence, in particular:

(3) $\mathrm{R} \in \mathrm{R} \leftrightarrow \mathrm{R} \notin \mathrm{R}$

which is a contradiction. In the same way by taking $\phi(y)$ to be the formula $y = y$, (NC) entails that there is a set which contains everything – the set of all sets in other words – and we can then derive Cantor's paradox.[43]

Zermelo replaces the principle of naïve comprehension the *axiom of separation*. This axiom states that for any property ϕ and any set x, there is a set y, which contains every element of x for which ϕ holds:

(AS) $\forall x \exists y(\forall z(z \in y \leftrightarrow (z \in x \,\&\, \phi(z))))$

[43] Although mathematicians had frequently assumed the existence of sets defined by abstraction, it would be misleading to say that they had implicitly assumed (NC). As far as I know, Frege is the only mathematician ever to have explicitly relied on a principle equivalent to (NC). Certainly Cantor did not. His definition of a set as 'a collection, gathered into a whole, of certain well-distinguished objects of our perception or our thought' [Cantor 1955, p. 85] is *not* an implicit statement of NC. This definition is in fact akin to Euclid's definition of a point as 'that which has no part', it is something which gives us a general idea of what we are talking about, but plays no role at all in the formal development of the theory. Indeed Cantor must have known that (NC) cannot be true in general, since he himself proved that there is no set of all sets and hence that the condition 'x is a set' does not determine a set.

This axiom does not by itself imply that there are any sets at all. Hence, as Zermelo points out, it cannot be used to deduce any of the known paradoxes: '...sets may never be *independently defined* by means of this axiom but must always be *separated* as subsets from sets already given.' [Zermelo 1908, p. 202]. For example if we already have a set A, then (AS) tells us that there must be a set consisting of all the elements of A which are not members of themselves. Call this set R. Then we have:

(1) $\forall z(z \in \mathrm{R} \leftrightarrow (z \in \mathrm{A} \,\&\, z \notin z))$

But we cannot get a contradiction from this as before. If we take z to be R itself, we obtain:

(2) $\mathrm{R} \in \mathrm{R} \leftrightarrow (\mathrm{R} \in \mathrm{A} \,\&\, \mathrm{R} \notin \mathrm{R})$

which is perfectly consistent and implies that R is not a member of itself or of A.

However, the axiom of separation does entail that there is no universal set, no set which contains every set. For if there were such a set U, there would be, by the axiom of separation, a subset of U which consists of all and only those elements of U which are not members of themselves. This subset would then be the set of all sets which are not members of themselves; a set which is a member of itself if and only if it is not a member of itself. This a contradiction, so there is no universal set. As noted above, this result – that there is no set of all sets – had already been established by Cantor in a different way.

Of course we do need to assume the existence of some sets in order to get the theory off the ground. Zermelo introduces special axioms which assert the existence of various sets: the axiom of elementary sets, which asserts the existence of the empty set $\varnothing$ and for all a and b, the sets $\{a\}$ and $\{a, b\}$, the power set axiom, which asserts the existence of the set of the set of all subsets of any set and the axiom of union, which asserts that given any family of sets A, there is a set which consists every element of every set in A. [ibid. pp. 202-3]. Although these axioms give us infinitely many sets, they are not enough to ensure the existence of any sets with infinitely many members. Hence Zermelo introduced the axiom of infinity, which states that there is a countably infinite set.[44]

The most controversial of Zermelo's axioms was the axiom of choice. This states that for any family of non-empty, mutually disjoint sets, there exists a set which contains one and only one element from each of those sets. Given such a family of sets, in other words, we can always choose one element from each set and form another set from the elements we choose [ibid. p. 204]. The axiom of choice was controversial at the time, since it implies the existence of certain sets without telling us how such sets can be independently defined or constructed in any particular case. How did Zermelo justify this axiom?

I mentioned that Zermelo had two motivations for axiomatizing set theory. The first was the avoidance of the paradoxes. The second motivation was his desire to prove a certain theorem, that every set can be well-ordered. Cantor had wanted to establish that any two sets can always be compared as to size, in the sense that for all sets A and B, either A $\succ$ B, A = B or A $\prec$ B. That, is given any two sets A and B either A can be put into one-one correspondence with a subset of B or B can be put into one-one correspondence with a subset of A. Cantor was unable to prove this result in general, but was able to show that it holds for well-ordered sets. Cantor conjectured that in fact, the result holds in full generality, since *every* set can be well-ordered. In 1900 Hilbert included the discovery of a proof of the well-ordering theorem in his list of the twenty three most important problems confronting the mathematical world.

[44] More precisely the axiom states that there is a set I which contains $\varnothing$ and for all a, if $a \in \mathrm{I}$ then $\{a\} \in \mathrm{I}$ [ibid. p. 104]. This formulation of the axiom stems from Dedekind, who had attempted to prove it by means of a rather curious argument [Dedekind 1888, §66].

In 1904 Zermelo was able to prove the well-ordering theorem, using the axiom of choice [Zermelo 1904, 1908b]. This establishes Cantor's conjecture and allow us to prove the principle of trichotomy for cardinal numbers:

(T) For all cardinal numbers a and b: either $a > b$, $a = b$ or $a < b$

Obviously, this principle is necessary if we are to have an adequate transfinite arithmetic and so a great deal of indirect justification accrued to the axiom of choice by Zermelo's proof that it implied this principle.[45]

Nonetheless, many mathematicians objected to Zermelo's proof. Peano for example, objected to the proof on the grounds that it was based on an unproved assumption. In response, Zermelo pointed out that since we cannot prove everything, every proof must begin from certain unproved principles. How are these principles to be justified? Zermelo's answer is clear: '[e]vidently by analysing the modes of inference that in the course of history have come to be recognized as valid and by pointing out that the principles are intuitively evident and necessary for science – considerations that can be urged equally well in favour of the disputed principle.' [Zermelo 1908b, p. 187].

Although Zermelo mentions self-evidence here, he regards it as too subjective to carry much weight. He continues '..the question that can be objectively decided, whether the principle is necessary for science, I should now like to submit to judgement by presenting a number of elementary and fundamental theorems and problems that, in my opinion, could not be dealt with at all without the principle of choice'. [ibid. pp. 187-8]. Zermelo then lists seven theorems from set-theory and analysis which depend on his new axiom. Hence for Zermelo, it is the power of the axiom of choice to prove important results

[45] In fact, it turns out that the trichotomy principle also implies the axiom of choice and so the two are in fact equivalent.

that provides the main source of its justification. Zermelo's view of the matter has since been amply vindicated; the axiom of choice is now a fundamental and indispensable principle in numerous and diverse branches of mathematics.[46]

In 1938 Gödel established that the axiom of choice is consistent relative to the other axioms of set-theory. Using the same technique, he was also able to establish that the continuum hypothesis is also consistent relative to the other axioms, which means that the continuum hypothesis is not disprovable from those axioms. [Gödel 1938,1939]. Gödel conjectured that continuum hypothesis was in fact undecidable on the basis of the standard axioms. In 1963 Paul Cohen proved Gödel right by showing how to construct models of the axioms in which the continuum hypothesis is false [Cohen 1963-4].[47] Hence neither (CH) nor its negation is provable from the standard axioms of set-theory. Since then, many other questions in set-theory have turned out to be undecidable on the basis of the standard axioms.[48]

Gödel had already canvassed the possibility of introducing new axioms which would allow us to settle such independent questions. But how would such axioms be justified? Gödel answers as follows:

> ....even disregarding the intrinsic necessity of some new axiom, and even in case it has no intrinsic necessity at all, a probable decision about its truth is possible also in another way, namely inductively by studying its 'success'. Success here means fruitfulness in consequences, in particular in 'verifiable' consequences, i.e consequences demonstrable without the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover... There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding

[46] See [Moore 1982] for a discussion of the history of the axiom of choice and a list of its important consequences and equivalents. For a more detailed analysis of the justification of the standard axioms for set theory (including the axioms of replacement and foundation which were added later by Skolem, Fraenkel, von Neumann and Zermelo) see [Maddy 1997, pp. 36-62].
[47] See [Maddy 1997, pp. 64-6] for further details.
[48] For some examples see [Maddy 1997, pp. 63-72].

> such powerful methods for solving problems...that, no matter whether or not they are intrinsically necessary, they would have to be accepted at least in the same sense as any well-established physical theory.
>
> [Gödel 1947, p. 477]

A great deal of contemporary work in set-theory consists of just his kind of investigation of new axiom candidates. Penelope Maddy has investigated the sort of arguments mathematicians use in justifying such axioms.[49] These come in several forms. Firstly, there are what Maddy calls *intrinsic* justifications. These appeal to various intuitions concerning the concept of a set; most frequently to the 'iterative conception' of the universe of sets.[50] Secondly, there are what Maddy calls *extrinsic* justifications. These appeal to the deductive power of the axioms in proving certain theorems. Maddy gives the following summary:

> The extrinsic evidence cited in previous sections came in a bewildering variety of forms, among them: (1) confirmation by instances (the implication of known lower-level results, ... principles known to be provable in ZFC);[51] (2) prediction (the implication of previously unknown lower level results ... later proved from ZFC alone); (3) providing new proofs of old theorems...; (4) unifying new results with old, so that the old results become special cases of the new...; (5) extending patterns begun in weaker theories; (6) providing powerful new ways of solving old problems...; (7) providing proofs of statements previously conjectured... (8) filling a gap in a previously conjectured "false, but natural proof"... (9) explanatory power ... (10) intertheoretic connections.
>
> [Maddy 1988b, p. 759]

Notice how justifications (1) to (4) appeal to the power of the new axioms to entail theorems that were either previously accepted or which were later shown to be acceptable independently of the new axioms. Justifications (5) to (10) also appeal to the deductive power of the new axioms, in generalizing or explaining the current body of results, or by

[49] In 'Believing the Axioms' [Maddy 1988ab]. See also [Maddy 1997, pp. 73-81].
[50] For a discussion of the iterative conception see [Boolos 1971] and [Parsons 1977].
[51] ZFC is Zermelo-Fraenkel set theory with the axiom of choice.

filling gaps in intuitively acceptable proofs. In contemporary set theory then, we find the same pattern that we have seen throughout our brief tour of the history of mathematics; axioms are justified in terms of their consequences, by their power to prove theorems.

3. Reflective Equilibrium

Axiomatization usually only occurs in mathematics once there is a fairly large body of results already available. In general, axiomatization is justified by the way in which it brings system and generality to a collection of accepted mathematical results. Particular axioms are justified if they yield all the theorems we already accept and none which we do not accept.

This seems to involve us in a vicious circle. Theorems are justified if they can be deduced from axioms and definitions, but axioms and definitions are justified if they entail the theorems. This situation is reminiscent of another apparently vicious circle noticed by Nelson Goodman in a different context:

> I have said that deductive inferences are justified by their conformity to valid general rules, and that general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.
>
> All this applies equally well to induction. An inductive inference, too, is justified by conformity to general rules, and a general rule by conformity to accepted inductive inferences.
>
> [Goodman 1983, p. 64]

According to Goodman, general rules of inference, both deductive and inductive are justified by showing that they yield all, or most of the particular inferences we accept and none or few which we do not. Particular inferences in turn are justified by showing that they conform to such general rules. The process of bringing our general rules and judgements about particular instances of them into agreement with each other has become known as the process of *reflective equilibrium.*

I argued in chapter one that the relationship between our logical intuitions and theories is indeed one of reflective equilibrium. A logical theory is justified to the extent that it provides us with a systematic account of the particular inferences we find intuitively acceptable or unacceptable and our intuitions about particular inferences may themselves be evaluated in terms of their conformity to our logical theories.

Consider, for example, classical first-order logic. Criticisms of this theory are of two main kinds; firstly, that it counts as valid particular inferences which are intuitively invalid, secondly, that it counts as invalid particular inferences which are intuitively valid[52]. As an example of the first kind, we have criticisms of various classically valid principles of deduction; that every statement follows from a contradiction or that every statement entails every logical truth. These principles license particular inferences which many find to be intuitively invalid, since the premises may have nothing in common with the conclusion.[53]

On the other hand, it can be argued that classical first-order logic fails to license particular inferences which we do want to accept. First-order logical entailment is *compact.* This means that if a statement follows from an infinite set of sentences, it follows from a

[52] A different kind of criticism of this theory concerns not so much the inferences which it licenses or fails to licence, as its descriptive or expressive power. The main rival theory here of course, is second-order logic. For a criticism of first order logic along these lines, see [Skolem 1922].
[53] See for example [Read 1988;1994] and [Haack 1978 pp. 197-203].

finite subset of those sentences. Consider an infinite set of sentences, each of which ascribes some property P to each natural number: {P0, P1, P2, P3, P4, P5 .... }. Now consider the sentence α: *every natural number has the property P.* If every sentence in the infinite set is true, it seems that α must also be true. So the infinite set of sentences seems to validly entail the sentence α. However, the infinite set of sentences does not classically entail the sentence α. For if it did, by compactness, some finite subset of the infinite set of sentences would entail A. But no *finite* subset of such sentences entails that *every* natural number has the property P, since it is possible that some, but not all natural numbers have the property P.[54] Here then, we have a case where our logical theory fails to license an inference which seems intuitively valid. Both types of criticism of deductive inference rules conform to Goodman's account. We can criticise a proposed rule either if it yields particular inferences we are unwilling to accept or if it fails to yield inferences which we do accept.

Similar remarks apply to inductive inference rules. Many such rules have been proposed and many criticisms of them have been put forward. Again these criticism have been of two main kinds. Either the proposed rule over-generates, yielding inferences which are intuitively unacceptable, or it under-generates, failing to yield inferences which are intuitively acceptable. We saw some examples in our discussion of the HD account of confirmation in the previous chapter. For example, (HD') implies that α confirms β for all α and β, provided that α and β are consistent. It therefore appears to license the inference from any statement to any other statement consistent with it. On the other hand, if the explanation criterion of confirmation is correct and if entailment is not necessary for

[54] See [Read 1994, pp. 42-49].

explanation, then there will be many intuitively valid inferences which (HD') *fails* to license.[55] Just as in the case of deductive logic, general principles of inductive inference are justified to the extent that they agree with our intuitions about the validity of particular inferences, while those intuitions themselves gain support from their agreement with our general principles.[56]

In exactly the same way, general principles in ethical and political theories (such as the principle of utility or the principle of equality) are justified by showing that they uphold our intuitions about the right or just action in particular cases. By the same token, such principles may be criticised on the grounds that they fail to conform to our intuitions about particular cases. But of course, our moral intuitions themselves are equally open to justification or criticism in the light of theoretical principles. The fact that a general principle conflicts with an intuition does not mean that the theory must be incorrect; it may be the intuition itself which needs to be abandoned.[57]

We can also see the process of reflective equilibrium at work in many other branches of philosophy, where we find for example, proposed analyses of concepts such as knowledge, causation, explanation and so on, defended and criticised on the grounds that

[55] See also [Putnam 1975h,1975i] for a criticism of Carnap's system of inductive logic along similar lines.
[56] Stich and Nisbett [1980] have argued that Goodman's account of the justification of inference rules cannot be correct, since many people actually reason according to patently invalid inference rules. A person who reasons according to the gambler's fallacy for example, adopts a general rule which conforms to the particular inferences they actually make and those inferences are in turn licensed by the rule. Hence, the rule is in reflective equilibrium with their inferential practice. Yet the rule is invalid and so reflective equilibrium is not sufficient to justify inductive inference rules. I believe that this criticism can in fact be met, although a full discussion of it would take us too far afield. It is worth pointing out that Stich and Nisbett do not abandon Goodman's account completely. Their solution to the problem they describe is to attempt to widen the notion of reflective equilibrium so that it is sensitive to the inferential practice of appropriate community authorities
[57] The method of reflective equilibrium in political and ethical theory is most commonly associated with the work of John Rawls. The phrase ***reflective equilibrium*** was in fact coined by Rawls who cites Goodman as the original source of the idea. See [Rawls 1971, pp. 20, 48-50].

they agree or fail to agree with our intuitive judgements concerning the application of these concepts in various contexts.

In a very general sense, the method of reflective equilibrium equally applies to the development of science. Here theories and laws play the role of general principles and observation and experiment correspond to intuitions. Our theories must agree with the results of observation, but observations themselves are not infallible and may be rejected, or at least treated with suspicion, if they are inconsistent with firmly established scientific theories. On the other hand, an observation which conflicts with theoretical principles may be accommodated by modification of auxiliary hypotheses or laws. The considerations which guide us in this process of bringing our theories and observations into agreement with each other are of course very complex and do not admit of a simple statement as a set of general principles of 'scientific method'. Nonetheless, the process does not seem essentially different to the rational assessment of theories by reflective equilibrium in other areas.

4. Reflective Equilibrium in Mathematics

What I want to suggest now is that Goodman's concept of reflective equilibrium is also applicable to the justification of our mathematical beliefs. Just as an inference rule is justified if it yields particular inferences we are willing to accept and none which we are unwilling to accept, while particular inferences are justified if they conform to such rules, I suggest that in mathematics, axioms and definitions are justified if they yield theorems we are willing to accept and none we are unwilling to accept and that theorems are justified if they can be deduced from such axioms or definitions.[58]

So for example, we can justify the associative law of addition by proving it from the axioms and definitions of number theory. But those axioms and definitions are themselves justified because we need to assume them in order to prove the accepted theorems of number theory, including the associative law itself. In the same way, Euclid provided a justification for Pythagoras's theorem by proving it from his axioms. But Euclid's axioms are justified by showing that they can be used to prove such accepted theorems of geometry. In general, a theorem may be justified by deducing it from axioms and definitions, while the axioms and definitions are justified by showing that they entail many statements we already accept and none that we would not accept

Why is this kind of justification of axioms, definitions and theorems not viciously circular? To answer this question, we need to take into account the point stressed by Lakatos and Kitcher, that mathematics is a subject with a history of change and evolution. This evolution does not consist merely in the fact that new results are proved as time goes on. It also consists in the fact that the same result can receive different justifications at different times. That is, to use Kitcher's terminology, the evolution of mathematics includes changes in the *set of accepted reasonings*.

Consider an axiom that is justified by showing that it entails a result we already accept. This sort of justification need not be circular because the result in question may

[58] Gödel seems to have arrived at a similar position. For him, some basic truths of mathematics can be known by intuition, a faculty akin to sense perception. More theoretical principles are justified in terms of their consequences. Moreover, the relationship between the first principles of a mathematical theory and the intuitive truths is something like the relationship of reflective equilibrium. [see Gödel, 1944, pp. 449-50 and 1947, pp. 477,483-5]. The account I develop below differs from Gödel's in replacing the somewhat vague notion of intuition with various kinds of non-deductive evidence, of a sort familiar from the natural sciences. For a more detailed discussion of Gödel's thinking on these matters see [Maddy 1997, pp. 89-94,172-6]. See also [Brown 1997, p. 168].

have been previously accepted on independent grounds. Proof from explicitly stated axioms is not the only source of justification in mathematics. We can have proofs, like Euclid's proof that every composite number is divisible by a prime number from implicit, unstated first principles. There are also what Kitcher calls unrigorous reasonings, such as the derivations of the derivatives of polynomial functions mentioned in chapter three.

A result which becomes accepted by appeal to arguments of this unrigorous kind, may later receive a rigorous justification by showing how it can be deduced from certain axioms. The axioms themselves are justified by showing that they entail such accepted results. But there is no vicious circle involved here, because the reason the result was initially accepted may be independent of the new reason provided by the derivation from axioms. What may happen then is that the old unrigorous justification of the result is removed from the set of accepted reasonings and replaced by the new more rigorous derivation from the axioms.

The following (artificial) example illustrates the sort of process I have in mind. Consider the following statement about the sum of the first n natural numbers.

$$1+2+3+4+\ldots.+n=\frac{1}{2}n(n+1)$$

This result was well known to the Pythagoreans, even though it is doubtful they had a rigorous proof of it.[59] It is possible to justify the claim in other ways however. Here is one way the Pythagoreans might have justified it.

Consider the diagram shown in figure 1. Each square in the diagram is a unit square, so the area of each is one unit. The total area of the figure is equal to the sum of the first six natural numbers. In this case then, we have $n = 6$. Now the area of the triangular part of the

[59] See [Kline 1972, p. 30].

figure, by elementary geometry is equal to half its base times its height. The base and height of the triangle are both equal to n, so the area of the triangle is equal to $\frac{1}{2}n^2$. To get the area of the whole figure we need to add the areas of the remaining shaded triangles. There are n of these and the area of each one of them is one half, so their total area is $\frac{1}{2}n$. Thus the total area of the triangle, equal to the sum of the first n natural numbers is:

$$1+2+3+4+\ldots+n \quad = \frac{1}{2}n^2+\frac{1}{2}n \quad = \frac{1}{2}n(n+1)$$

This is not of course, a rigorous proof. We inferred from a particular diagram, involving the first six natural numbers that the result holds for all natural numbers. Nonetheless, it seems that we have given quite a good intuitive argument for the result.

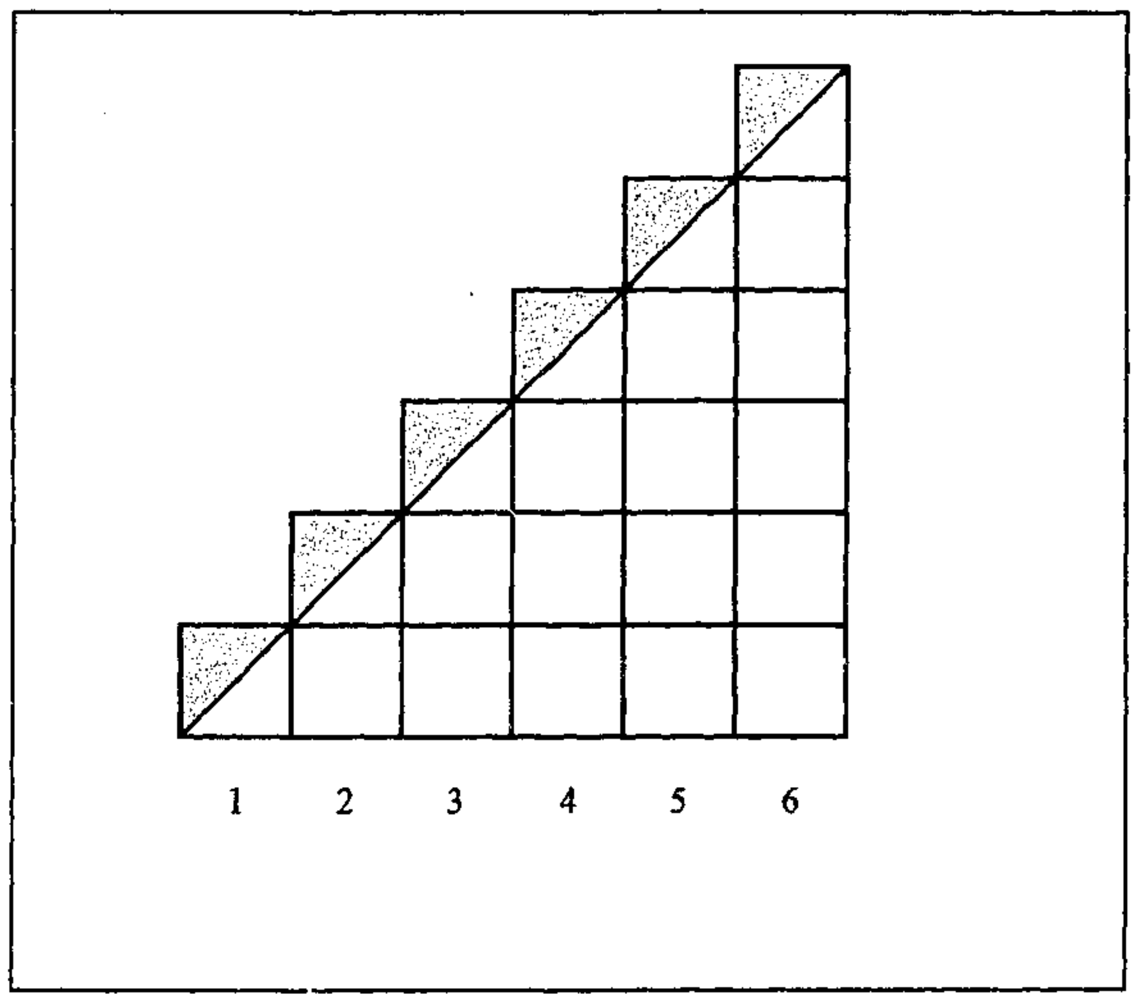


Figure 1

Suppose the Pythagoreans came to accept this result on something like these grounds. Given that the result has become accepted, we can imagine introducing an axiom that is justified by showing that it entails the result. As an example, suppose we introduce the principle of mathematical induction as an axiom. We can then use this axiom to prove our result more rigorously. First we show that the result holds for the case $n = 0$. But in this case:

$$1+2+3+4+\ldots+n = 0$$

and:

$$\frac{1}{2}n(n+1) \quad = \frac{1}{2}\cdot 0(0+1)$$
$$= 0$$

Now we show that if the result holds for any number n, it also holds for $n+1$. So we assume that:

(1) $\quad 1+2+3+4+\ldots+n = \frac{1}{2}n(n+1)$

and we need to show that (1) entails:

(2) $\quad 1+2+3+4+\ldots+n+(n+1) \; = \frac{1}{2}(n+1)((n+1)+1)$
$$= \frac{1}{2}(n+1)(n+2)$$

But if we add $(n+1)$ to both sides of (1) we obtain:

$$\begin{aligned}1+2+3+4+\ldots+n+(n+1) &= \frac{1}{2}n(n+1)+(n+1)\\ &= (n+1)\left(\frac{1}{2}n+1\right)\\ &= (n+1)\left(\frac{1}{2}(n+2)\right)\\ &= \frac{1}{2}(n+1)(n+2)\end{aligned}$$

which establishes (2) as required.

We began with a perceptual, intuitive justification of a statement. We then introduced a new axiom, the principle of mathematical induction. This principle was justified by showing that it could be used to deduce our previously accepted statement. At the same time, that statement itself received a new and more rigorous justification. The old intuitive argument may then come to be replaced by the derivation from the new axiom.

This example is not of course meant to be historically accurate. It is simply meant to be illustrative of the process of reflective equilibrium at work in mathematics. It shows that this process need not involve a vicious circle, since an axiom may be justified in terms of a theorem which was accepted for some other, perhaps less rigorous reason.

This process is well illustrated by the examples from the history of mathematics considered in section two. Sometimes the evidence for an axiom or definition is that it entails a result that has already been proved independently. An example would be the evidence for Cantor's set theory provided by his proof of the existence of transcendental numbers; a result which had already been proved, along different lines, by Liouville and others. In the same way, the axiom of separation gains some support by showing that it entails that there is no universal set, a result which had already been proved by Cantor via the power-set argument. Such independent proofs may be more or less rigorous of course; Hilbert's axioms for example, are supported by showing how they entail theorems of Euclidean and non-Euclidean geometry which had already been proved less rigorously. Similarly, Cauchy's definition of the derivative is supported by showing how it can be used to derive the standard rules of differentiation, rules which had previously been established only been means of a various unrigorous arguments.

Sometimes, however, the evidence for an axiom is that it entails an accepted result that has never been proved before, not even by means of an unrigorous argument. An examples would be Zermelo's proof of the well-ordering theorem, which provided evidence for the axiom of choice. In the same way, evidence for the Dedekind-Peano axioms is provided by the derivation of the such principles as the associative law of addition - principles which no one had ever *proved* before. Similarly, Dedekind's theory of cuts is supported by the derivation of the limit existence theorems which had defeated Cauchy.

But for such derivations to count as evidence for the axioms and definitions, there must be some independent reason for accepting these results. That is, there must be some independent *evidence* for them. That evidence is not provided by a proof, for here we are considering results which are being proved for the first time. This points to the existence of a kind of evidence in mathematics other than that provided by rigorous or not so rigorous deductive argument – *non-deductive evidence*, in fact.[60] In chapter four, I suggested that there can be empirical evidence for mathematical statements. On the other hand, we also find *quasi-empirical* evidence in mathematics; evidence which is derived from other *mathematical* statements (rather than from empirical ones) by means of such familiar

---

[60] Of course, the evidence for *axioms* must be of this kind, since axioms are not proved. So we have already seen one kind of non-deductive evidence in mathematics; an axiom or definition can be supported by showing that it implies an accepted result. The present point is just that those accepted results may themselves be supported by various kinds of non-deductive evidence.

patterns as induction and analogy.[61] In the case of such principles as the associative laws of addition and multiplication, we can say that although these had no *proof* prior to their derivation from the axioms of number theory, they were certainly well supported by a large body of this kind of evidence. Non-deductive mathematical evidence comes in a wide variety of different forms, some of which I would now like to discuss in more detail. Let us begin by taking a closer look at *induction* in mathematics.

5. NON-DEDUCTIVE EVIDENCE

5.1 INDUCTION

Here we find evidence for a universal statement derived from the verification of a number of its instances. Inductive evidence of this kind is exceedingly common in mathematics.[62] Number theory provides a rich source examples. I[illegible]40, Fermat investigated numbers of the form:

$$2^{2^n}+1$$

When $n = 0$, we have $2^{2^0}+1 = 3$ which is a prime number. When $n = 1$, we have $2^2+1=5$ and 5 is also prime. For $n = 2$, we obtain $2^4+1= 17$ and 17 is prime. The next two numbers in the sequence are 257 and 65537 and Fermat verified that these are also prime numbers. On the basis of this evidence, Fermat conjecture that *every* number of the form $2^{2^n}+1$ is prime. In fact, Fermat's conjecture turned out to be false. Euler, in 1732,

[61] The term *quasi-empirical* was introduced by Hilary Putnam. See [Putnam 1975c, p. 62]

[62] In fact, it is much easier to find real examples of induction in mathematics than it is to find such examples in the natural sciences.

showed that the very next number in the sequence $2^{32}+1$ is not prime, since it is divisible by 641.[63]

Fermat did claim to have proved his famous 'last theorem', that the equation:

$$x^n = y^n + z^n$$

has no positive integer solutions when $n \geq 3$, but as is well known, his proof was never found and it now seems unlikely that he had a proof. Evidence for Fermat's conjecture however, came in the form of verification of particular instances. Fermat claimed to have proved it for the case $n = 4$. Euler later gave a proof for this case and for the case $n = 3$.[64] These results already show that Fermat's conjecture is true for an *infinite* number of values of n. Since it is true for $n = 3$, for example, it must also be true for any multiple of 3. For suppose the conjecture is false for some n which is a multiple of 3. Then there would be integers x, y, z and m such that:

$$x^{3m} = y^{3m} + z^{3m}$$

But then we would have:

$$(x^m)^3 = (y^m)^3 + (z^m)^3$$

and the conjecture would fail for $n = 3$. The case of $n = 4$, on the other hand, shows that Fermat's last theorem is true for any n which is not divisible by an odd prime.[65] But it is easy to show that if Fermat's conjecture is true for every prime number greater than two,

[63] In fact, $2^{32}+1 = 4294967297 = 641 \cdot 6700417$. See also [Kline 1972, pp. 277,609, Polya 1954a p. 9, Scarlau and Opolka 1985, p. 9].

[64] See [Kline 1972, pp. 276-7,609, Devlin 1988, pp. 187-9].

[65] For suppose that Fermat's last theorem is false for some n which is not divisible by a prime $p > 2$. Since every number is divisible by some prime number, n must be divisible by 2. Hence n is either 2 or a power of 2. Since n is greater than 2, n must be of the form $n = 4m$. But by the same reasoning as that given above for the case of $n = 3$, the case of $n = 4$ establishes that Fermat's equation has no solutions for any n which is a multiple of 4. Hence Fermat's conjecture is true for any n which is not divisible by an odd prime.

then it is true for any n which *is* divisible by an odd prime.[66] Hence, if we could establish that Fermat's equation has no solutions when n is an odd prime number, we would have a proof of the conjecture. Mathematicians therefore set about showing that Fermat's conjecture is true for particular prime values of n. In 1825, Dirichlet and Legendre independently proved the case of $n = 5$ and in 1839 Lamé gave a proof for $n = 7$.[67]

The first general result along these lines however, was given by Sophie Germain in 1823. She proved that the equation:

$$x^p + y^p = z^p$$

has no solutions when p is a prime number which does not divide the product xyz and where $2p + 1$ is also prime.[68] Many similar results, for various kinds of prime exponent were obtained in the nineteenth century. Ernst Kummer, for example, established that there are no solutions when p is a so called *regular* prime[69]. The only irregular primes less than 100 are 37, 59 and 67 and Kummer was able to show that there are no solutions to Fermat's equation for these values of n. This establishes that Fermat's Last Theorem is true for all values of n up to 100.[70]

With the invention of the computer in the twentieth century, mathematicians were able to use this kind of technique to establish that Fermat's Last Theorem held up to larger

[66] Suppose n is divisible by an odd prime. Then $n = ap$ for some a and some prime $p > 2$. If Fermat's conjecture was false for such an n we would have integers, x, y and z such that $x^{ap} + y^{ap} = z^{ap}$. But then we would have $(x^a)^p + (y^a)^p = (z^a)^p$ and the conjecture would be false for some prime > 2. Hence if Fermat's conjecture is true for every odd prime, it is true for every number divisible by an odd prime.

[67] An accessible account of these proofs (and the proofs for $n = 3$ and $n = 4$)can be found in [Devlin 1988, pp. 182-92]. See also [Kline 1972, pp. 609, Stewart 1987, p. 27, Burton 1998, p. 490].

[68] See [Devlin 1988, p. 196-7, Silverman 1996, p. 22].

[69] The following definition is equivalent to (but simpler than) the one Kummer gave. A prime p is *regular* if and only if it does not divide the numerators of the ***Bernoulli numbers*** B_2, B_4, ... B_{p-3} where B_k is the coefficient of x^k in the infinite series expansion of $x/(e^x - 1)$. See [Devlin 1988, p. 194].

[70] See [Kline 1972, p. 818-820, Devlin 1988, pp. 192-5, Stewart 1987, pp. 29-30, Burton 1998, pp. 491-2].

and larger values of n. By this means S. Wagstaff in 1978 established that there are no solutions to Fermat's equation for all values of n up to 125,000.[71]

Euler published many papers in which he stated hypotheses which he could not prove, along with as much inductive evidence for them as he could find. A beautiful example is the short piece 'Discovery of a most extraordinary law of numbers concerning the sum of their divisors', translated in full by Polya [1954a, pp. 91-8]. Euler begins his paper as follows:

> Till now the mathematicians tried in vain to discover some order in the sequence of the prime numbers and we have every reason to believe that there is some mystery which the human mind will never penetrate. To convince oneself, one has only to glance at the tables of the primes...and one perceives that there is no order and no rule... I am myself certainly far from this goal, but I just happened to discover an extremely strange law governing the sums of the divisors of the integers which, at the first glance appear just as irregular as the sequence of the primes, and which, in a certain sense, comprise even the latter. This law...is in my opinion, so much more remarkable as it is of such a nature that we can be assured of its truth without giving it a perfect demonstration. Nevertheless, I shall present such evidence for it as might be regarded as almost equivalent to a rigorous demonstration.
>
> [ibid. p. 91]

Euler then introduces the sigma function, $\sigma(n)$ equal to the sum of the divisors of n. So for example $\sigma(2) = 1 + 2$, $\sigma(6) = 1 + 2 + 3 + 6 = 15$ and so on. Clearly, if p is a prime number then $\sigma(p) = p + 1$. Euler gives a table in which he calculates the values of $\sigma(n)$ for all values of n from 1 to 99. The first few values are 1,**3**,**4**,7,**6**,12,**8**,15,13,18.... Here the values of $\sigma(n)$ for which n is prime are marked in bold. Euler remarks that if we examine this sequence '...we are almost driven to despair...The irregularity of the primes is so deeply involved in it that we must think it impossible to disentangle any law governing this

[71] See [Wagstaff 1978]. For further information on the history of Fermat's Last Theorem see [Singh 1998, Kline 1972, pp. 276-7,609,818-20, Stewart 1987, pp. 24-34, Burton 1998, pp. 489-93, Devlin 1988, pp. 177-200]. See also [Silverman 1996, Chapter 37] for an account of Wile's final proof.

sequence.' [ibid. p. 93]. Nonetheless, Euler claims that in fact each number in this sequence can be computed from earlier ones by means of a certain rule he has discovered. The rule is:

$$\sigma(n)= \sigma(n-1) + \sigma(n-2) - \sigma(n-5) - \sigma(n-7) + \sigma(n-12) + \sigma(n-15) - \sigma(n-22) - \sigma(n-26) + \ldots$$

To apply this rule we add up the terms of this sequence, ignoring those for which the argument of the sigma function is less than zero and substituting the value n if the argument is equal to zero. So for example, according to Euler's rule, the value of $\sigma(7)$ is given by:

$$\sigma(7) = \sigma(6) + \sigma(5) - \sigma(2) - \sigma(0) = 12 + 6 - 3 - 7 = 8$$

which is correct. The difference between consecutive terms in the sequence of numbers 1, 2, 5, 7, 12, 15, 22 ... which we have to subtract from n is given by the sequence 1,3,2,5,3,7,4,9,5.... where we have, alternatively, all the integers 1, 2, 3, 4,.... and the odd numbers 3, 5, 7, 9,.... Hence, as Euler remarks '...it is not difficult to apply the formula to any given particular case, and so anybody can satisfy himself of its truth by as many examples as he may wish to develop. And since I must admit that I am not in a position to give it a rigorous demonstration, I will justify it by a sufficiently large number of examples.' [ibid. p. 93]. He then shows that his rule gives the correct answer for the first twenty values of n. He notes however, that these cases only make use of the first six numbers in the sequence 1, 2, 5, 7, 12, 15 ... and so he also shows that his rule gives the correct answer for the case of $\sigma(100)$ (which uses the first 16 of these numbers) and $\sigma(301)$ (which uses the first 28 of them). These examples, says Euler, 'will undoubtedly dispel any qualms which we might have had about the truth of my formula.' [ibid. p. 95].

Euler does not end there however. He goes on to explain how he discovered this strange rule and the same time, cites further evidence for it. He begins with the infinite product:

(1) $$(1-x)(1-x^2)(1-x^3)(1-x^4)(1-x^5)(1-x^6).....$$

By multiplying out larger and larger numbers of terms in this sequence, Euler came to the conclusion that (1) expands out to:

(2) $$1-x-x^2+x^5+x^7-x^{12}-x^{15}+x^{22}+x^{26}-\ldots$$

Notice that the exponents of x in (2) are exactly the same as the numbers used in Euler's rule for the sigma function. Euler then shows that indeed, on the assumption that the infinite product in (1) is equal to the infinite sequence given by (2), we can prove that his rule is correct. But he cannot prove this assumption – the only evidence he has for it is inductive evidence, found by multiplying out finite numbers of the terms in (1). Nonetheless, '...each of us can convince himself of this truth by performing the multiplication as far as he may wish; and it seems impossible that the law which has been discovered to hold for 20 terms, for example, would not be observed in the terms that follow.' [ibid. 96]. Euler concludes that '[t]his reasoning, although still very far from a perfect demonstration, will certainly lift some doubts about the most extraordinary law that I explained here.' [ibid. p. 98].

In a letter to Euler of 1742, Christian Goldbach made a conjecture which is equivalent to the following claim:

Every even number greater than two is the sum of two primes

Euler was convinced the conjecture was true, although he could not prove it.[72] Again, we can acquire some evidence for the conjecture by checking particular cases:

$$\begin{gathered} 4 = 2 + 2 \\ 6 = 3 + 3 \\ 8 = 3 + 5 \\ 10 = 3 + 7 \\ 12 = 5 + 7 \end{gathered}$$

and so on. Goldbach's conjecture has now been confirmed in this way for every even number up to $4 \cdot 10^{14}$. [See Richstein 2000]. No one has been able to prove that it holds for all even numbers however, although it is known that every even number is the sum of not more than six primes and that 'almost all' even numbers are the sum of two primes.[73]

A related hypothesis is the *twin-primes conjecture*:

There are infinitely many prime numbers p such that p + 2 is also prime.

for which there is also a great deal of inductive evidence of a similar kind. Although the conjecture remains unproved, computers have been used to establish the existence of larger and larger twin-primes. A recent record was set by Giovanni La Barbera[74], who found the pair $1693965 \cdot 2^{66443} \pm 1$. In addition, Chen Jing-run has shown that something close to the twin-primes conjecture is true: there are infinitely many primes p such that $p+2$ is either prime or the product of two primes.[75]

By examining tables of prime numbers, Gauss, Legendre and others noticed that the number of primes less than n is roughly equal to $n/\log(n)$. More precisely, as n gets larger

[72] See [Kline 1972, p. 610].

[73] If P(n) is any property of integers and #P(N) is the number of integers n less than N which satisfy P(n), then if the ratio #P(N)/N approaches 1 as N approaches infinity, P is said to hold for 'almost all' integers. For more information on Goldbach's conjecture see [James 1949].

[74] See [Caldwell 2000, p. 1]

[75] Chen Jing-run also proved a similar result on the Goldbach conjecture: every even number $n >$ ...s the sum of a prime and a number which is either prime or the product of two primes. See [Silverman 1996, 78-9].

and larger, the ratio of the number of primes less than n to $n/\log(n)$ gets closer and closer to 1 (see Table 1). These observations provided inductive evidence for the *prime number theorem*:

$$\lim_{n \to \infty} \frac{\pi(n)}{n / \log(n)} = 1$$

where $\pi(n)$ gives the number of primes less than or equal to n. The prime number theorem was proved first by Hadamard and Poussin in 1896 using methods from complex analysis, and then again in 1948 by Paul Erdös using only 'elementary' methods.[76]

n	$\pi(n)$	$n/\log(n)$	$\frac{\pi(n)}{n/\log(n)}$
10	4	4.34	0.921
100	25	21.71	1.151
1000	168	144.76	1.161
10^4	1229	1085.74	1.132
10^6	78498	72382.41	1.084

TABLE 1
Inductive evidence for the prime number theorem

In their proofs Hadamard and Poussin had made use of the so called *zeta function*:

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \ldots$$

first introduced by Euler in 1737 and then generalized by Riemann in 1859 to the case where s is an arbitrary complex number. The solution to the problem of finding a proof of

[76] See [Kline 1972, pp. 830-2, Stewart 1987, p. 125-7]. Table 1 has been compiled from data in [Silverman 1996, p. 77].

the prime number theorem depends on the nature of the values of s for which $\zeta(s) = 0$, the *zeros* of the zeta function. The *Riemann hypothesis* is that:

All the complex zeros of the zeta function have the form ½ + iy.

Riemann established that this hypothesis implies the prime number theorem, but offered no proof of it[77]. It remains one of the most important outstanding problems in mathematics, having far reaching implications for many questions in number theory, especially those concerning the distribution of the primes. One kind of evidence for the Riemann hypothesis is purely inductive; one can calculate the zeros of the zeta function and check that they all have real part ½. Computer calculations have been used to establish that Riemann's hypothesis holds for the first 1.5 billion zeros of the zeta function.[78]

This kind of evidence is not considered very compelling however. For example, R. F Churchhouse and I. J Good have argued that there are reasons to believe that the first zero of the zeta function with real part not equal to ½, if there is one, might have an imaginary part as large as $10^{10,000}$. If so, then it might never be practically possible to find this zero by calculation.[79] The calculations have also revealed some 'near misses', values of the zeta function which are very close to zero, but with real part not equal to ½. Various kinds of more compelling non-deductive evidence for the Riemann hypothesis are mentioned below.[80]

Inductive evidence is not confined to number theory. For example, noticing that a second degree equation has two roots, a third degree equation has three roots, a fourth

[77] In [Riemann 1859]. The independent proofs of the prime number theorem due to Hadamard and Poussin do not depend on the Riemann hypothesis of course. The prime number theorem turns out to be equivalent to the claim that none of the complex zeros of the zeta function have real part ≥ 1.
[78] See [van de Lune *et al* 1986].
[79] See [Davis and Hersh 1981, p. 364].
[80] For further details of the history and evidence for Riemann's hypothesis see [Edwards 1974, Kline 1972, p. 831, Stewart 1987, pp. 43,125-7, Kolata 1974, Franklin 1987, pp. 4-8, Brown 1999, pp. 166-7].

degree equation has four roots and so on, Albert Girard inferred that an *n*th degree polynomial equation has *n* roots, if we include what he called the 'impossible' (complex) roots.[81]

We have already seen one example for geometry; Euler's evidence for his conjecture on polyhedra was obtained by verifying that the formula V– E + F = 2 holds for the regular polyhedra, as well as a various prisms and pyramids.[82] Another example is the *isoperimetric theorem*, stated by Descartes:

Of all plane figures of equal area, the circle has the least perimeter.

We can gain some inductive evidence for this by making comparing the perimeter of a circle of area one unit with the perimeters of various other shapes with the same area (see Table 2). On the basis of this kind of evidence, Descartes was convinced: '[i]n order to show by enumeration that the perimeter of a circle is less than that of any other figure of the same area, we do not need a complete survey of all the possible figures, but it suffices to prove this for a few particular figures whence we can conclude the same thing, by induction, for all the other figures'.[83]

[81] See [Kline 1972, p. 270]
[82] In [Euler 1750]. See also [Polya 1954a, pp. 35-43].
[83] From Descartes' 'Rules for the Direction of the Mind' [1628] cited in Polya [1954a, p. 168] from which Table 2 has been compiled. For a discussion of the inductive evidence for this and related conjectures, see [Polya 1954a, pp. 168-189].

Figure	*Perimeter*
Circle	3.55
Square	4.00
Rectangle 3:2	4.08
Semicircle	4.10
Rectangle 2:1	4.24
Equilateral Triangle	4.56
Isosceles right triangle	4.84

TABLE 2
Inductive evidence for the isoperimetric theorem

Finally, we might mention Euler's famous result, mentioned in chapter three, that:

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\ldots=\frac{\pi^2}{6}$$

and hence $\zeta(2) = \pi^2/6$. Euler's argument for this result, discussed below, was based on analogy, rather than induction. However, Euler verified the result by calculating sums of finite segments of this infinite series and comparing them to the value of $\pi^2/6$ (see Table 3). This technique works inductively, we acquire evidence that the infinite sum is correct by checking that a sufficiently large finite segment of the series adds up to a value which agrees sufficiently with our answer.[84]

[84] For further examples of inductive evidence in number theory, geometry and analysis, see [Polya 1954a, pp. 43-52,59-70,76,7,79-83, 1954b, p. 3-4]. See also [Stewart 1987, p. 31,127] for a discussion of the inductive evidence for the Mordell conjecture in algebraic geometry and the Bieberbach conjecture in geometric function theory. On the role of inductive evidence in the classification of the finite simple groups see [Franklin 1987, pp. 9-13].

n	$\sum_1^n \frac{1}{n^2}$	$\frac{\pi^2}{6}-\sum_1^n \frac{1}{n^2}$
10	1.549	0.095
20	1.596	0.048
30	1.6121	0.032
40	1.6202	0.024
50	1.6251	0.019
100	1.6349	0.009
150	1.6382	0.006

TABLE 3
Inductive evidence for Euler's theorem

All these examples seem to instantiate a common pattern. But what is that pattern? The obvious thing to say is that a statement of universal form is confirmed by any of its instances. Using the notation of the previous chapter, we might express this idea with the following schema:

(I) $\alpha \equiv \forall x(\Phi x \rightarrow \Psi x)\ \&\ \beta \equiv (\Phi a\ \&\ \Psi a) \Rightarrow \beta\mathbf{C}\alpha$

from which we can derive:

$$(Fa_1\ \&\ Ga_1)\ \mathbf{C}\ \forall x(Fx \rightarrow Gx)$$

$$(Fa_2\ \&\ Ga_2)\ \mathbf{C}\ \forall x(Fx \rightarrow Gx)$$

$$\ldots$$

$$(Fa_n\ \&\ Ga_n)\ \mathbf{C}\ \forall x(Fx \rightarrow Gx)$$

Unfortunately, as is well known, the obvious thing to say is incorrect. For example, as instance of (I) we have:

$$\alpha \equiv \forall x(\sim Gx \rightarrow \sim Fx)\ \&\ \beta \equiv (\sim Ga\ \&\ \sim Fa) \Rightarrow \beta\mathbf{C}\alpha$$

But then if we take $\alpha = \forall x(Fx \rightarrow Gx)$ and $\beta = (\sim Fa\ \&\ \sim Ga)$ we can infer that:

(RP) $(\sim Fa\ \&\ \sim Ga)\mathbf{C}\forall x(Fx \rightarrow Gx)$

which means that an object which is neither F nor G confirms the universal statement that all Fs are Gs. This is the well known *paradox of the ravens*, mentioned in the previous chapter. To see how it applies in the mathematical case, let Fx be the predicate 'x is an even number greater than two' and let Gx be the predicate 'x is the sum of two prime numbers'. Then $\forall x(\mathrm{F}x \rightarrow \mathrm{G}x)$ is Goldbach's conjecture – every even number greater than two is the sum of two primes. (I) entails that this is confirmed by the instances F4 & G4, F6 & G6, F8 & G8 and so on. But we have just shown that according to (I), Goldbach's conjecture is also confirmed by any statement of the form ~Fa & ~Ga, that is, by any object a which is not an even number greater than 2 and not the sum of two primes. The irrational number $\sqrt{2}$ is an example of such an object; it is neither an even number, nor is it the sum of two primes. This observation provides evidence for Goldbach's conjecture according to (I). Any other irrational number would do just as well of course, as would an infinite number of other mathematical objects (the set of all real numbers, the unit circle, Kepler's 'urchin' and so on). Worse, the mug of coffee on my desk is an object which is not an even number and not the sum of two primes, so again, this observation also confirms Goldbach's conjecture according to (I). Similar statements about desks, tables and chairs would also confirm it, for the same reason. The problem with (I) is that it entails that all sorts of statements which appear to be completely irrelevant to Goldbach's conjecture actually provide evidence for it. But that is not all. Notice that just as $\sqrt{2}$ is not the sum of two primes, it is not the sum of *three* primes either. So $\sqrt{2}$ is also an instance of the generalization 'all numbers not the sum of three primes are not even numbers greater than two'. So according to (I) $\sqrt{2}$ confirms not only Goldbach's conjecture, but also the conjecture that all even numbers greater than two are the sum of three primes. In exactly the same way, it will confirm every

hypothesis of the form 'every even number greater than two are the sum of n primes', for any n. In general, (I) entails that any object which is not a natural number will provide us with evidence for an infinite number of incompatible hypotheses.

Another very famous problem with the induction schema (I) is Goodman's 'grue' paradox.[85] Let us consider a mathematical example. Above, I mentioned that Goldbach's conjecture has been confirmed for all even numbers up to $4 \cdot 10^{14}$. Let Fx and Gx be as before and let Hx be the predicate 'x is the sum of three primes'. Suppose we then introduce the predicate GRUE(x) defined by:

$$\mathrm{GRUE}(x) \leftrightarrow (\mathrm{G}x \,\&\, x \leq 4 \cdot 10^{14}) \vee (\mathrm{H}x \,\&\, x > 4 \cdot 10^{14})$$

That is, a number is GRUE if and only if it is either less than or equal to $4 \cdot 10^{14}$ and the sum of two primes or greater than $4 \cdot 10^{14}$ and the sum of three primes. It is not hard to show that if $n \leq 4 \cdot 10^{14}$ then:

(1) $$\mathrm{F}n \,\&\, \mathrm{G}n \equiv \mathrm{F}n \,\&\, \mathrm{GRUE}(n)$$

That is, if $n \leq 4 \cdot 10^{14}$, then the statement that n is F and G is equivalent to the statement that n is F and GRUE.[86] Now, as an instance of (I) we have:

(2) $$\alpha \equiv \forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x)) \,\&\, \beta \equiv (\mathrm{F}n \,\&\, \mathrm{GRUE}(n)) \Rightarrow \beta \mathbf{C} \alpha$$

so if we let $\alpha = \forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x))$ and $\beta_n = (\mathrm{F}n \,\&\, \mathrm{G}n)$, then we have, for all $n \leq 4 \cdot 10^{14}$:

$$(\mathrm{F}n \,\&\, \mathrm{G}n)\mathbf{C}\forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x))$$

and in particular:

[85] See [Goodman 1983, pp. 81-83].

[86] Suppose $n \leq 4 \cdot 10^{14}$. If (Fn & Gn) then we have Gn & $n \leq 4 \cdot 10^{14}$. By $\vee$-introduction, we have: (Gn & $n \leq 4 \cdot 10^{14}$) $\vee$ (Hn & $n > 4 \cdot 10^{14}$). Hence n is GRUE and so we have (Fn & GRUE(n)). Conversely, if (Fn & GRUE(n)) then we have Fn and (Gn & $n \leq 4 \cdot 10^{14}$) $\vee$ (Hn & $n > 4 \cdot 10^{14}$). But since, by our supposition., $n \leq 4 \cdot 10^{14}$, the second disjunct is false and we can infer that (Gn & $n \leq 4 \cdot 10^{14}$). Hence, n is G and so we have (Fn & Gn).

$$\mathrm{F}4\ \&\ \mathrm{G}4\ \mathbf{C}\forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x))$$

$$\mathrm{F}6\ \&\ \mathrm{G}6\ \mathbf{C}\forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x))$$

$$\mathrm{F}8\ \&\ \mathrm{G}8\ \mathbf{C}\forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x))$$

$$\ldots$$

$$\mathrm{F}(4\cdot 10^{14})\ \&\ \mathrm{G}(4\cdot 10^{14})\ \mathbf{C}\ \forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x))$$

Hence, according to (I), all the inductive evidence we have for Goldbach's conjecture is equally confirming evidence for the conjecture that every even number greater than two is GRUE. But Goldbach's conjecture implies that every even number *greater* than $4\cdot 10^{14}$ is the sum of two primes, while the latter conjecture implies that every even number greater than $4\cdot 10^{14}$ is the sum of *three* primes. In other words, our evidence confirms mutually incompatible hypotheses. Of course, we can replace Hx in the definition of GRUE(x) with any thing we like ('x is the sum of four primes', 'x is not divisible by any prime number', 'x is the largest city in Australia' and so on) and then (I) will imply that the inductive evidence which supports Goldbach's conjecture also provides us with evidence that every even number greater than $4\cdot 10^{14}$ has any of these properties too.

These difficulties should not be taken to imply that a mathematical statement can never be confirmed by inductive evidence of the kind we have been discussing. Rather, they show that our initial statement of the relevant rule (I) needs to be reformulated. I shall say something about how we might approach this problem in the final section.

5.2. ANALOGY

Another important kind of non-deductive evidence in mathematics is based on analogy. Here we obtain evidence that an object (or class of objects) has a certain property by arguing that it is analogous to some other object (or class of objects) known to have the property. The most famous example is Euler's derivation of sum of the series:

$$1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\frac{1}{25}+\ldots$$

Euler found the sum of this series by means of an argument based on an analogy with polynomial equations.[87] To explain Euler's reasoning we need to review some facts about such equations. A polynomial equation has the general form:

(1) $$a_0+a_1x+a_2x^2+a_3x^3+\ldots+a_nx^n=0$$

An n-th degree polynomial equation has n roots; r_1, r_2, r_3, …, r_n. Furthermore, it can be shown that any such polynomial can be factorised in the following way:[88]

(2) $$a_0+a_1x+a_2x^2+a_3x^3+\ldots+a_nx^n=a_0\left[\left(1-\frac{x}{r_1}\right)\left(1-\frac{x}{r_2}\right)\left(1-\frac{x}{r_3}\right)\ldots\left(1-\frac{x}{r_n}\right)\right]$$

That is, we can express the polynomial as a product of n linear factors, with a factor corresponding to each root. Now suppose our polynomial equation is of the form:

(3) $$a_0+a_2x^2+a_4x^4+a_6x^6+\ldots\ldots+a_{2n}x^{2n}=0$$

Such an equation of degree $2n$ will have $2n$ roots of the form: $r_1, -r_1, r_2, -r_2, \ldots r_n, -r_n$. By (2) the polynomial on the left hand side of this equation can be written as a product of $2n$ linear factors:

[87] Euler's argument has been widely discussed. See for example [Polya 1954a, pp. 17-21,30-4, Putnam 1975c, pp. 67-8, Steiner 1975, Kitcher 1984, pp. 196-7 Kline 1972, pp. 448-9, Franklin 1987, p. 4, Brown 1999, pp. 168-70]. Euler's original argument and his response to critics such as Daniel Bernoulli can be found in his *Opera Omnia* [1911-36, series 1, vol. 8, p. 168 and vol. 14, pp. 73-86, 138-55].

[88] Provided only that $a_0 \neq 0$.

$$a_0+a_2x^2+a_4x^4+\dots+a_{2n}x^{2n}=a_0\left[\left(1-\frac{x}{r_1}\right)\left(1-\frac{x}{(-r_1)}\right)\left(1-\frac{x}{r_2}\right)\left(1-\frac{x}{(-r_2)}\right)\dots\left(1-\frac{x}{r_n}\right)\left(1-\frac{x}{(-r_n)}\right)\right]$$
$$=a_0\left[\left(1-\frac{x}{r_1}\right)\left(1+\frac{x}{r_1}\right)\left(1-\frac{x}{r_2}\right)\left(1+\frac{x}{r_2}\right)\dots\left(1-\frac{x}{r_n}\right)\left(1+\frac{x}{r_n}\right)\right]$$
$$=a_0\left[\left(1-\frac{x^2}{r_1^2}\right)\left(1-\frac{x^2}{r_2^2}\right)\dots\left(1-\frac{x^2}{r_n^2}\right)\right]$$

Writing out the last line again, we have:

(4) $$a_0+a_2x^2+a_4x^4+\dots+a_{2n}x^{2n}=a_0\left[\left(1-\frac{x^2}{r_1^2}\right)\left(1-\frac{x^2}{r_2^2}\right)\dots\left(1-\frac{x^2}{r_n^2}\right)\right]$$

So, for example a fourth degree equation of the form:

$$a_4x^4+a_2x^2+a_0=0$$

has four roots, r_1, $-r_1$, r_2, $-r_2$. Hence, by (4) we have:

(5) $$a_4x^4+a_2x^2+a_0=a_0\left[\left(1-\frac{x^2}{r_1^2}\right)\left(1-\frac{x^2}{r_2^2}\right)\right]$$

The reader may care to verify, for example, that the equation $x^4-13x^2+36=0$, which has the four roots 2, –2, 3, –3 can indeed be written in the form $36\left[\left(1-\frac{x^2}{4}\right)\left(1-\frac{x^2}{9}\right)\right]$. Now, multiplying out the right hand side of equation (5), we obtain:

$$a_4x^4+a_2x^2+a_0 \quad =a_0\left[\left(1-\frac{x^2}{r_1^2}\right)\left(1-\frac{x^2}{r_2^2}\right)\right]$$
$$=a_0\left[1-\frac{x^2}{r_1^2}-\frac{x^2}{r_2^2}+\frac{x^4}{r_1^2r_2^2}\right]$$
$$=a_0-a_0\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)x^2+a_0\left(\frac{1}{r_1^2r_2^2}\right)x^4$$

Comparing the coefficients of x^2 on the left and right hand sides of this equation, we obtain the relation:

$$-a_2=a_0\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)$$

In fact, this relationship between the roots and coefficients is quite general. By comparing the coefficients of x^2 on the left and right hand side of (4), we can deduce that for an arbitrary polynomial of the form (3), we have:

(6) $$-a_2=a_0\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}+\frac{1}{r_3^2}+\dots+\frac{1}{r_n^2}\right)$$

We can now explain Euler's argument. Euler begins with the power series expansion of the function sin(x):

$$\sin(x)=\frac{x}{1}-\frac{x^3}{3\cdot2\cdot1}+\frac{x^5}{5\cdot4\cdot3\cdot2\cdot1}-\frac{x^7}{7\cdot6\cdots3\cdot2\cdot1}+\dots$$

Euler proposes to think of this as a *polynomial of infinite degree*. Support for this analogy comes from the fact the equation has an infinite number of roots, corresponding to the values of x for which sin(x) = 0, namely, 0, π, -π, 2π, -2π, 3π, -3π and so on. Dividing the equation through by x, we obtain:

$$\frac{\sin(x)}{x}=1-\frac{x^2}{3\cdot2\cdot1}+\frac{x^4}{5\cdot4\cdot3\cdot2\cdot1}-\frac{x^6}{7\cdot6\cdots3\cdot2\cdot1}+\dots$$

In the same way, we can think of this as an infinite polynomial with roots π, -π, 2π, -2π, 3π, -3π....Notice that the expression on the right hand side of this equation has the form of (3) above. In this case we have $a_0=1$, $a_2=-\frac{1}{3\cdot2\cdot1}$ and so on. By analogy with (4) then, Euler concludes that the infinite polynomial on the right can be expressed as an infinite product of factors, with one factor corresponding to each double root:

(7) $$\frac{\sin(x)}{x} = a_0\left[\left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{(2\pi)^2}\right)\left(1-\frac{x^2}{(3\pi)^2}\right)\left(1-\frac{x^2}{(4\pi)^2}\right)......\right]$$
$$=\left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{4\pi^2}\right)\left(1-\frac{x^2}{9\pi^2}\right)\left(1-\frac{x^2}{16\pi^2}\right)......$$

But then by equation (6) – the relation between the roots and coefficient of x^2 - we have:

$$-a_2 = a_0\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}+\frac{1}{r_3^2}+\frac{1}{r_4^2}+.....\right)$$
$$\frac{1}{2\cdot 3} = \left(\frac{1}{\pi^2}+\frac{1}{(2\pi)^2}+\frac{1}{(3\pi)^2}+\frac{1}{(4\pi)^2}+......\right)$$
$$\frac{1}{6} = \left(\frac{1}{\pi^2}+\frac{1}{4\pi^2}+\frac{1}{9\pi^2}+\frac{1}{16\pi^2}+......\right)$$

Multiplying through by π^2, we obtain the conclusion:

(8) $$\frac{\pi^2}{6} = 1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\frac{1}{25}....$$

Euler's argument is ingenious. It does not amount to a proof however, since Euler could give no rigorous argument for the identity (7) – his evidence is based entirely on the analogy with the case of (4) for ordinary finite polynomials. Nonetheless, as mentioned above, Euler was able to verify his solution by calculation. Furthermore, he used the same technique to derive further results, such as:

$$\frac{\pi^2}{8} = \frac{1}{1^2}+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+....$$
$$\frac{\pi^4}{90} = \frac{1}{1^4}+\frac{1}{2^4}+\frac{1}{3^4}+\frac{1}{5^4}+....$$

which could also be checked by calculation. Euler was also able to use his technique to derive a result obtained independently[89] by Leibniz:

(9) $$\frac{\pi}{4} = 1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+....$$

Hence Euler's technique could be used to establish a known result. Euler comments: '[f]or our method, which may appear to some as not reliable enough, a great confirmation here comes to light. Therefore, we should not doubt at all of the other things which are derived by the same method'.[90] Euler eventually found a proof of the sum given by (8) along more usual lines, which was accepted as a rigorous demonstration. Again, this independent verification of a consequence of Euler's method provided further evidence for the analogy and the results obtained from it.

Newton also made use of analogical evidence in his work on the calculus. In order to differentiate or integrate certain functions, Newton would use infinite series representations and then differentiate or integrate term by term.[91] For example, when faced with the problem of integrating a function such as:

$$y = \frac{1}{1+x^2}$$

Newton starts by expanding the function as a power series:

(10) $$\frac{1}{1+x^2} = 1-x^2+x^4-x^6+.....$$

and then integrates term by term, obtaining:

$$\int\frac{1}{1+x^2}dx = x-\frac{x^3}{3}+\frac{x^5}{5}-\frac{x^7}{7}+....$$

[89] Though not of course rigorously, at least by modern standards. See chapter three, section five.
[90] Cited in [Polya 1954a, p. 21].
[91] See [Kline 1972, pp. 360-1, Kitcher 1984, pp. 233-4].

But how did Newton obtain the infinite series representations of functions like that in (10)? Here he made use of the *binomial theorem*:

$$\text{(BT)} \qquad (1+x)^n = 1 + nx + \frac{n(n-1)}{2\cdot 1}x^2 + \frac{n(n-1)(n-2)}{3\cdot 2\cdot 1}x^3 + ...$$

which had long be known to hold when n is any positive whole number.[92] So for example, to evaluate $(1+x^2)^2$ we can set $n = 2$ and substitute x^2 for x in (BT). We then obtain:

$$\begin{aligned}(1+x^2)^2 &= 1 + 2(x^2) + \frac{2(2-1)}{2\cdot 1}(x^2)^2 + \frac{2(2-1)(2-2)}{3\cdot 2\cdot 1}(x^2)^3 + ...\\ &= 1 + 2x^2 + (2-1)x^4 + 0 + ...\\ &= 1 + 2x^2 + x^4\end{aligned}$$

which we can check is correct by multiplying out the right hand side. Now we can write our difficult function in the form:

$$y = \frac{1}{1+x^2} = (1+x^2)^{-1}$$

By analogy with (BT) then, we ought to have:

$$\begin{aligned}\frac{1}{1+x^2} &= (1+x^2)^{-1}\\ &= 1 + (-1)(x^2) + \frac{(-1)(-1-1)}{2\cdot 1}(x^2)^2 + \frac{(-1)(-1-1)(-1-2)}{3\cdot 2\cdot 1}(x^2)^3 + ...\\ &= 1 - x^2 + \frac{(2)}{2\cdot 1}x^4 + \frac{(-6)}{3\cdot 2\cdot 1}x^6 + ...\\ &= 1 - x^2 + x^4 - x^6 + ...\end{aligned}$$

which is the series in (10). In the same way, if we want to differentiate or integrate the function:

$$y = \sqrt{(1+x)} = (1+x)^{1/2}$$

[92] See [Kline 1972, pp. 272-4] for a discussion of the early history of the binomial theorem.

Then again, by analogy with (BT) we have:

$$\begin{aligned}(1+x)^{1/2} &= 1 + \tfrac{1}{2}x + \frac{\tfrac{1}{2}\left(\tfrac{1}{2}-1\right)}{2\cdot 1}x^2 + \frac{\tfrac{1}{2}\left(\tfrac{1}{2}-1\right)\left(\tfrac{1}{2}-2\right)}{3\cdot 2\cdot 1}x^3 + ...\\ &= 1 + \frac{1}{2}x + \left(-\frac{1}{4}\right)\frac{1}{2\cdot 1}x^2 + \left(\frac{3}{8}\right)\frac{1}{3\cdot 2\cdot 1}x^3 + ...\\ &= 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 + ...\end{aligned}$$

and we can then differentiate or integrate the function term by term. Newton became convinced that (BT) is true when n is any negative number or fraction, but gave no proof. His evidence was based partly on the analogy with the case when n is a positive whole number and partly on inductive verification of the result for special cases. For example, he multiplied out many terms in the product of the above series with itself and verified that the result does appear to yield $(1+x)$.[93]

Mathematicians often appeal to the proof (or disproof) of an analogous theorem as evidence for (or against) an unproved conjecture. For example, one of the reasons many mathematicians are suspicious of the inductive evidence for the Riemann hypothesis is that a related conjecture concerning the rate of growth of the Möbius function (see next section) which was supported by a similarly huge amount of inductive evidence, was eventually proved false.[94] On the other hand, in 1947 André Weil considered various generalizations of Riemann's zeta function and conjectured that the analogue of the Riemann hypothesis holds for these zeta functions. The Weil conjectures were eventually proved in 1975 and

[93] See [Kline 1972, p. 273, Kitcher 194, p. 234]. Euler also appears to have accepted the generalized binomial theorem on similar grounds; by analogy with the case where n is a positive whole number and by verification of the series obtained in particular cases by multiplication. See [Euler 1770, pp. 120-6].
[94] This was the so called *Mertens Conjecture*. See [Devlin 1988, pp. 216-221, Stewart 1984, p. 126].

many mathematicians have taken this as providing strong evidence that the Riemann hypothesis itself is also true.[95]

5.3 Statistics

In more recent years, a fruitful source of evidence for conjectures in number theory has been based on an analogy between certain properties of the natural numbers and the properties of random sequences studied by probability theory. Many of the special functions studied in number theory, such as $\pi(n)$ – the number of primes less than n and Euler's function $\varphi(n)$ – the number of integers less than n which have no common factor with it, yield sequences which appear to be random in the statistical sense. Hence probability theory can be used to cast some light on the properties of such functions. This analogy has developed into a completely rigorous branch of mathematics, known as *probabilistic number theory*.[96] The analogy can also be used to provide various plausibility arguments for unproved conjectures. For example, the twin-primes conjecture has been argued for by noting that the occurrence of twin-primes in the sequence of primes appears to be random. This suggests that if we pick a number x at random, the probability that it is prime is independent of the probability that $x + 2$ is prime. Now the prime number theorem implies that if n is large enough then the number of primes less than n is about $n/\log(n)$. Hence, if we choose a number x at random between 0 and n, the *probability* that x is prime will be approximately:

$$\frac{n/\log(n)}{n} = \frac{1}{\log(n)}$$

[95] See [Edwards 1974, p. 298. Franklin 1987, p. 6]. Weil's conjectures were themselves supported by various analogous theorems in topology, before they were p:oved. See [Stewart 1984, pp. 32-3] for more information.
[96] See [Hardy and Wright 1979, Kac 1959, Kubilius 1964].

Thus, if the probability that x is prime and the probability that $x + 2$ is prime are independent of each other, the probability that *both* x and $x + 2$ are prime, will be the product of these independent probabilities:

$$\frac{1}{\log(n)} \cdot \frac{1}{\log(n+2)} = \frac{1}{\log(n)\log(n+2)}$$

But this function approaches $1/\log(n)^2$ asymptotically. That is:

$$\lim_{n\to\infty} \frac{\dfrac{1}{\log(n)\log(n+2)}}{1/\log(n)^2} = 1$$

and therefore the probability that both x and $x + 2$ are prime is approximately equal to $1/\log(n)^2$. In other words, the expected number of twin-primes between 0 and n is approximately:

$$\frac{n}{\log(n)^2}$$

Since this fraction gets infinitely large as n goes to infinity, we can conclude that there are an infinite number of twin-primes in all.[97]

A similar argument has been suggested in support of the Riemann hypothesis.[98] Suppose we toss a fair coin N times. For large values of N we would expect to find roughly equal numbers of heads and tails, although in any particular trial would not expect the numbers to be exactly equal. In fact, it can be shown that on the average, the difference between the number of heads and the number of tails is $\sqrt{N}$. We can express this fact precisely as follows:

[97] For further details of this ment see [Caldwell 2000]. Caldwell also shows how we can construct similar plausibility arguments for Goldbach's conjecture and many other conjectures concerning primes. See also [Hardy and Wright 1979, 22.20, Putnam 1975c, p. 68-9, Davis and Hersh 1981, p. 364].
[98] This argument was first suggested by Denjoy [see Edwards 1974, pp. 268-9] and by Churchhouse and Good [1968]. See also [Franklin 1987, pp. 6-7, Davis and Hersh 1981, pp. 364-7].

(11) For all N and any $\varepsilon > 0$, $|\, \mathrm{NUM}_N(\mathrm{H}) - \mathrm{NUM}_N(\mathrm{T})\,| < N^{1/2+\varepsilon}$

where $\mathrm{NUM}_N(\mathrm{H})$ is the number of heads and $\mathrm{NUM}_N(\mathrm{T})$ the number of tails obtained in *N* tosses. Consider any random sequence of heads and tails obtained in a trial of *N* tosses. For example:

H H T T H H T T H H T H H T H T H H H H II T H H T H H H H T

Suppose we introduce a function *s*(*x*) which takes the value 1 if there is a H at position *x* in this sequence and takes the value –1 if we find a T at position *x* in the sequence. The function *s*(*x*) then replaces every H with a 1 and every T with a –1 so that we get the sequence:

1 1 -1 -1 1 1 -1 -1 1 1 -1 1 1 -1 1 -1 1 1 1 1 1 -1 1 1 -1 1 1 1 1 -1

Now suppose that we add up all the values of *s*(*x*) from 1 to *N*. Obviously the result will be equal to the number of heads in the original sequence minus the number of tails. So if we introduce the function:

$$sum(N) = \sum_{x=1}^{x=N} s(x)$$

Then we have:

$$sum(N) = \mathrm{NUM}_N(\mathrm{H}) - \mathrm{NUM}_N(\mathrm{T})$$

Which allows us to rewrite (11) as:

(12) For all N and any $\varepsilon > 0$, $|\, sum(N)\,| < N^{1/2+\varepsilon}$

and this will hold for any function *sum*(*N*) which is defined in this way in terms of some random sequence of two equiprobable independent events.

Now consider the *Möbius function* $\mu(n)$, defined to be 0,1, or -1, if in the prime factorization of *n* there is a repeated factor, an even number of factors or an odd number of

factors respectively.[99] If we look at the first sixteen values of $\mu(n)$, we discover the sequence:

1 –1 –1 0 –1 1 –1 0 0 1 –1 0 –1 1 1 0

This sequence of 1s, 0s and –1s is apparently random (see also table four). If we calculate a large number of further values of the function $\mu(n)$, we find this random pattern continued; 1s and –1s occurring with roughly equal probability. Furthermore, it seems likely that the successive values of $\mu(n)$ are independent of each other, since knowing the value of $\mu(n)$ does not seem to give us any information about the value of $\mu(n+1)$. Hence, if we define the function M(*N*) as follows:

$$M(N) = \sum_{n=1}^{n=N} \mu(n)$$

Then by analogy with (12), we ought to have:

(13) For any N and any $\varepsilon > 0$, $|\, M(N)\,| < N^{1/2+\varepsilon}$

But (13) is in fact equivalent to the Riemann hypothesis! In this way, the statistical analogy provides us with at least a plausible argument for the truth of Riemann's conjecture.

5.4. Verification of a Consequence

Perhaps the most common form of non-deductive evidence for a mathematical statement is obtained by the verification of one or more of its consequences. We have already seen this kind of evidence at work in the justification of axioms and definitions. Verification of a consequence can equally provide evidence for lower level conjectures.

[99] We treat $n = 1$ as a special case; by definition $\mu(1) = 1$.

n	Factors	$\mu(n)$
1	1	1
2	2	−1
3	3	−1
4	2·2	0
5	5	−1
6	2·3	1
7	7	−1
8	2·2·2	0
9	3·3	0
10	2·5	1
11	11	−1
12	2·2·3	0
13	13	−1
14	2·7	1
15	3·5	1
16	2·2·2·2	0

TABLE 4
Values of the Möbius Function

We have come across several examples already. Recall that Euler argued for his conjecture on polyhedra by showing that it implies a known result; that there are exactly five regular polyhedra. In the same way, identities such as (7) above, which Euler obtained by means of the analogy with polynomials, are supported by showing that they imply known results such as Leibniz' identity (9) $\frac{\pi}{4}=1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\ldots$. Further support for (7) comes from Euler's independent proof of the consequence (8) $\frac{\pi^2}{6}=1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\frac{1}{25}\ldots$ We also saw how Newton obtained evidence for the generalized binomial theorem by verification of some of its consequences, by checking for example that it implies the right answer for the expansion of $(1+x)^{\frac{1}{2}}$. Similar evidence can be found for the Riemann hypothesis; we have already mentioned the fact that it implies the prime number theorem

for example, which was proved independently by Hadamard and Poussin.[100] Another consequence of the Riemann hypothesis was established by Hardy in 1914; that there are infinitely many zeros of the zeta function of the form $\frac{1}{2} + iy$.[101]

We can also think of induction and analogy as special cases of this form of non-deductive evidence. A conjecture of the form 'all *F*s are *G*s' entails that if *a* is *F*, then *a* is *G*. So given some object *a* which is *F*, establishing that *a* is also *G* verifies this consequence of the conjecture and hence confirms it. On the other hand, since the generalized binomial theorem implies the restricted binomial theorem for positive integer exponents, we can think of Newton's analogical evidence for the former as being derived from the latter known consequence of it. In the same way, since Euler's identities for 'infinite polynomials' imply the known results for finite polynomials as a special case, and so we can think of his analogy as providing evidence for those identities via the verification these consequences of them.

I would like to mention one final example, which played an important role in the acceptance of complex numbers by the mathematical community. By examining the series expansions of e^x, $\sin x$ and $\cos x$, Euler obtained the fundamental identity relating the complex numbers to the exponential and trigonometric functions:

(14) $$e^{i\theta} = \cos\theta + i\sin\theta$$

from which, if we let $\theta = \pi$, we obtain the celebrated result:

$$e^{i\pi} + 1 = 0$$

Now from (14) we can derive the identity known as *De Moivre's theorem*:

[100] If the Riemann hypothesis is true, then we get a much better estimate of how much $\pi(n)$ differs from $n/\log(n)$ than that obtained by these independent proofs. [see Stewart 1987, p. 126, Kline 1972, pp. 830-2].
[101] See [Kline 1972, p. 831, Edwards 174, pp. 226-9].

(15) $$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$$

This was an important result for the developing theory of complex numbers because it has implications concerning properties of the trigonometric functions which do not involve complex numbers at all; properties which can therefore be verified independently. In particular, we can use De Moivre's theorem to obtain the standard expressions for the sine or cosine of any given multiple of an angle. For example, if we set $n = 2$ in (15) we have:

(16) $$(\cos\theta + i\sin\theta)^2 = \cos 2\theta + i\sin 2\theta$$

Multiplying out the left hand side, we have:

(17) $$\begin{aligned}(\cos\theta + i\sin\theta)^2 &= (\cos\theta)^2 + 2\cos\theta i\sin\theta + (i\sin\theta)^2\\ &= \cos^2\theta + 2i\sin\theta\cos\theta - \sin^2\theta\\ &= (\cos^2\theta - \sin^2\theta) + i(2\sin\theta\cos\theta)\end{aligned}$$

By comparing the real and imaginary parts of (16) and (17), we obtain the identities:

$$\begin{aligned}\cos 2\theta &= \cos^2\theta - \sin^2\theta\\ \sin 2\theta &= 2\sin\theta\cos\theta\end{aligned}$$

which can of course be proved without the use of complex numbers, by elementary trigonometry. By taking $n = 3, 4, 5, \ldots$ in (15) and comparing real and imaginary parts of the resulting expressions, we can also derive the known trigonometric identities for $\sin 3\theta$, $\cos 3\theta$, $\sin 4\theta$, $\cos 4\theta$ and so on. The verification of these consequences not only provided some support for De Moivre's theorem itself, but also indirectly provided evidence for the existence of complex numbers, by showing how they can be used to explain, in a systematic way, various known properties of the trigonometric functions.[102]

[102] This example is discussed by Frege in the *Grundlagen* in the context of his critique of formalism [Frege 1884, §97]. See also [Dummett 1991, pp. 178;284]. For more information on Euler's identity, De Moivre's theorem and the history of complex numbers see [Stewart 1987, p. 119-130, Kline 1972, pp. 253-4;408-11;628-32, Crossley 1980, pp. 131-232 and Nagel 1979].

5.5 VISUALIZATION

The final source of non-deductive evidence I would like to discuss is that provided by visualization. Here we acquire evidence for a statement by the use of the visual imagination, perhaps aided with a picture or diagram. I described one example in section four above, the derivation of the sum of the first n natural numbers by means of the diagram in figure 1.[103] This technique was frequently used by the Pythagoreans; they represented numbers as collections of dots or pebbles and classified them according to the various shapes that could be made with the dots.[104] So for example, they called the numbers 1, 3, 6, 10 ... *triangular* numbers because the dots can be arranged into the shape of a triangle, while the numbers 1, 4, 9, 16, ... were *square* numbers because the dots could be arranged into squares (see figure 2).

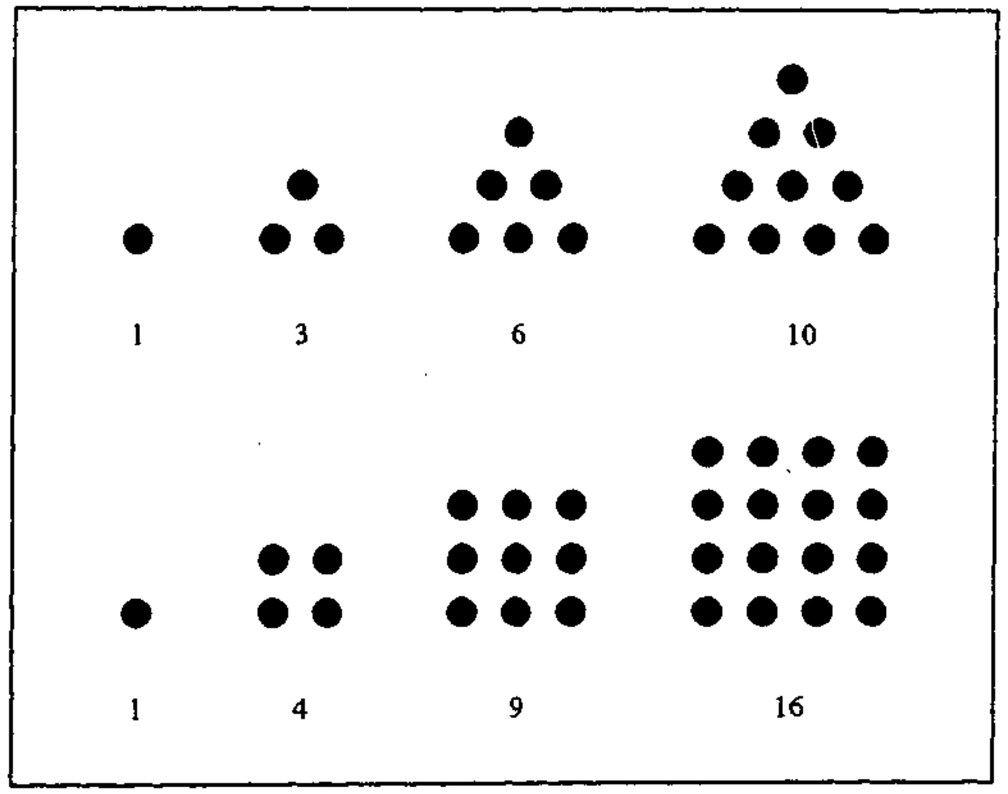


Figure 2
Triangular and Square Numbers

[103] See [Giaquinto 1993, pp. 386-7] for an alternative visual demonstration of this theorem.
[104] See [Kline 1972, pp. 28-34].

From these geometrical arrangements of dots, the Pythagoreans were able to obtain some simple properties of the natural numbers. For example, looking at figure 3a, it is obvious that the triangular numbers form the sequence 1, 1 + 2, (1 + 2) + 3, (1 + 2 + 3) + 4 ... which is to say that the *n*th triangular number represents the sum of the first *n* natural numbers. Looking at figure 3b, on the other hand, it becomes clear that the square numbers form the sequence, 1, 1 + 3, (1 + 3) + 5, (1 + 3 + 5) + 7... which is to say that the *n*th square number represents the sum of the first *n* odd numbers. That is:

$$1+3+5+7+\ldots+(2n-1)=n^2$$

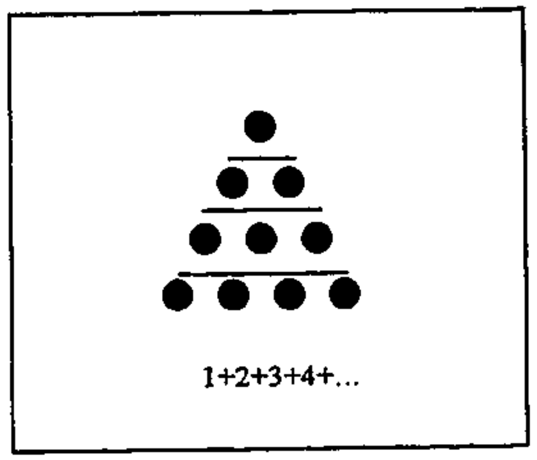


Figure 3a

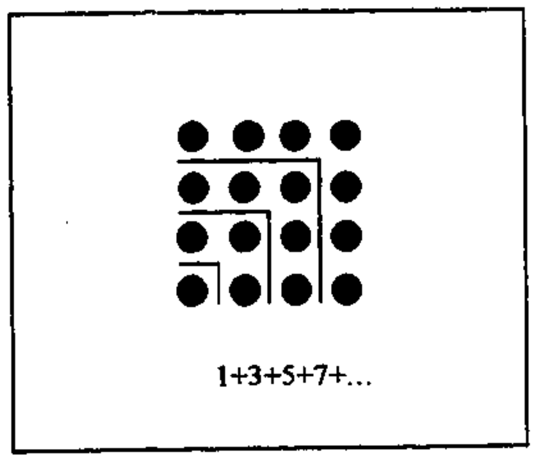


Figure 3b

My favourite example of visual reasoning in elementary arithmetic is the following.[105] Why is that two even numbers make an even number, an even and an odd number make an odd number and two even numbers make an even number? If we imagine the even numbers as rectangular arrays of squares, two squares high and any number of squares long, then the odd numbers will be represented by rectangular arrays with one extra square on the end. If we then think of addition as the result of putting two such arrays of

[105] I discovered this example in [Sawyer 1964, pp. 8-11].

squares together, we can see immediately why these facts about the addition of even and odd numbers must hold (see figure 4).[106]

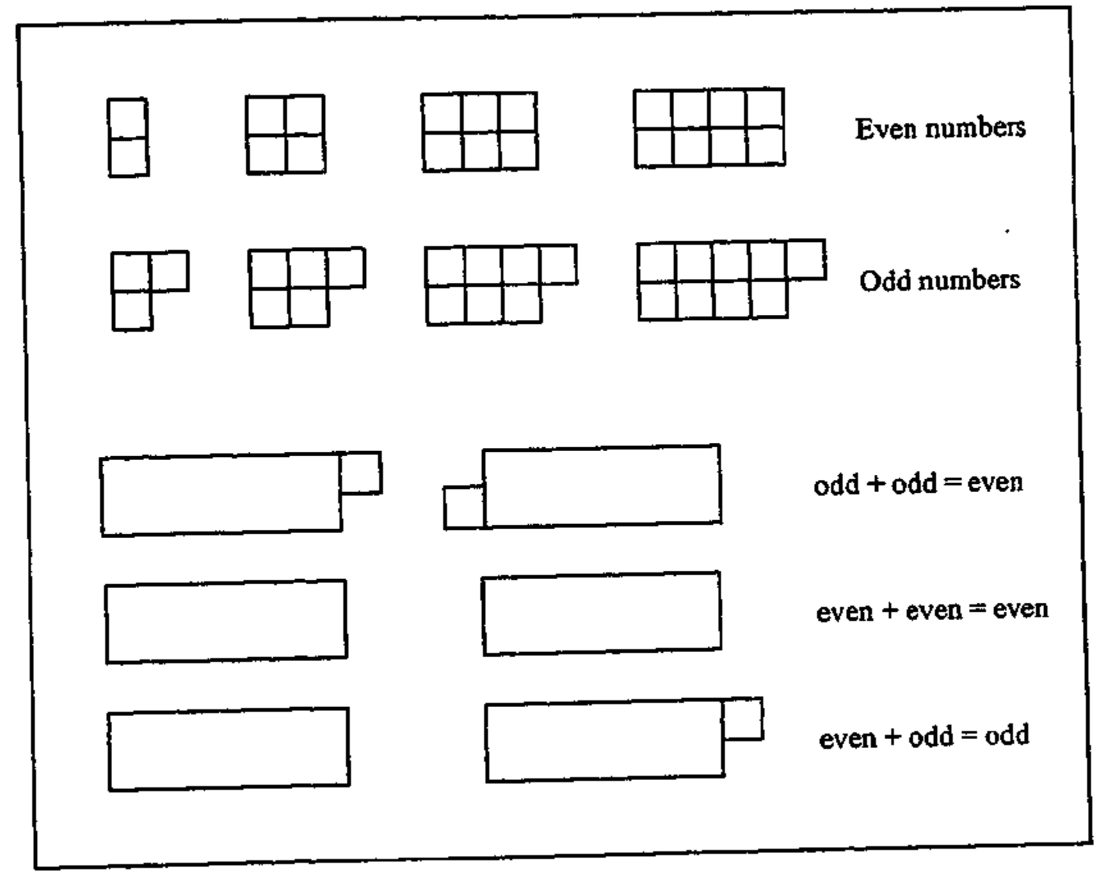


Figure 4
Even and odd numbers

The use of visual reasoning is most readily apparent in geometry of course. Imagine a triangle oriented so that the longest side is the base *b*. Now draw in the perpendicular from the upper vertex to the base; this represents the height *h* of the triangle. If you then embed the triangle in a rectangle of the same base and height, we get a figure with twice the area of the original triangle. Since the area of the rectangle is $b\times h$, the area of the triangle must be $\frac{1}{2}(b\times h)$. (see figure 5).

As another example, consider the following claim:

[106] It might be objected that this kind of reasoning cannot provide us with knowledge of general theorems in arithmetic, because any picture or visual image will be of some particular finite number of dots or squares and will therefore not apply to *all* numbers. For a response to this objection see [Giaquinto 1993].

It is impossible to completely cover a circle (so that there are no gaps) with a finite number of non-overlapping smaller circles.

Most people find this claim obviously true on the basis of visual intuition, although it might be quite difficult to prove (see figure 6).[107]

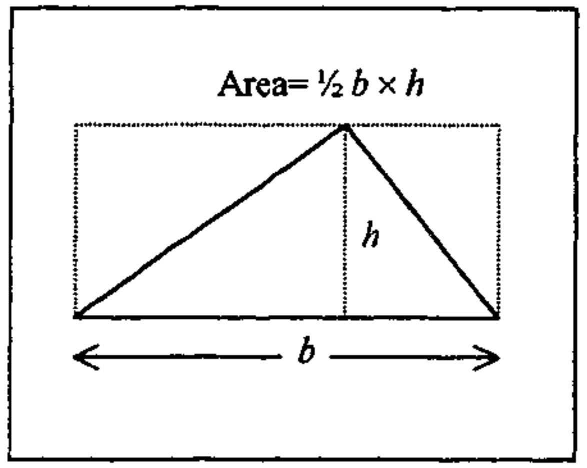


Figure 5
The area of a triangle

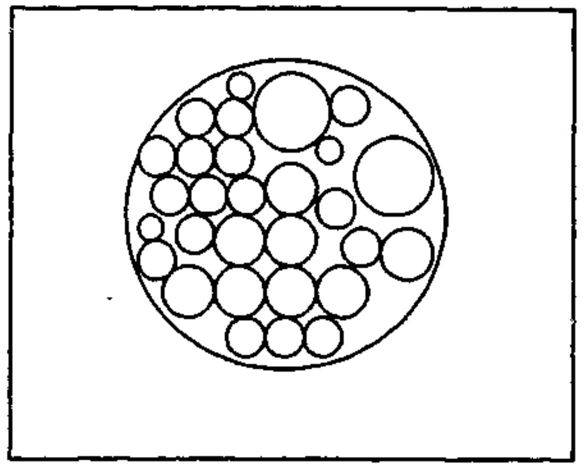


Figure 6
Covering a circle

We have already mentioned an example of visual reasoning in analysis; the 'geometric' argument for the intermediate value theorem. If we imagine any continuous curve which is negative at point a and positive at point b, it just seems obvious that the curve must cross the x-axis at some point (figure 7). Here is another example (figure 8). Imagine a unit square. Now divide it vertically in half, making two rectangles. Now imagine shading in the left hand rectangle. The area of the shaded rectangle is obviously equally to half the original area. Now take the unshaded rectangle and divide that in half with a horizontal line, making two squares. The area of each of these squares is obviously equal to one quarter of the original area. If we shade in the lower square, then the total area of the shaded region is equal to ½ + ¼.

[107] This example was described to me by Marcus Giaquinto, in conversation. See also [Giaquinto 1992].

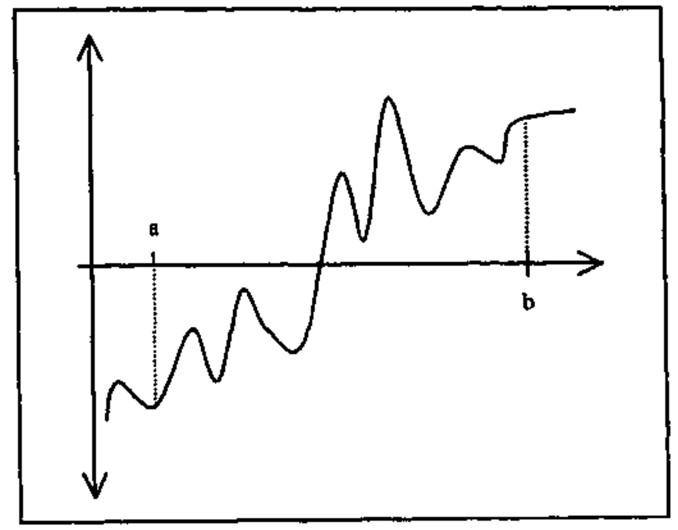


Figure 7
The intermediate value theorem

Now consider the upper unshaded square and divide it in half with a vertical line, making two smaller rectangles again. If we shade in the right hand rectangle, we will now have a total shaded area equal to ½ + ¼ + ⅛. We can imagine repeating this process *ad infinitum*. Clearly the area of the shaded region increases slightly at each stage, and covers more and more of the original square, but equally clearly, the shaded region can never exceed the area of the original square. From this we can infer that:

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots=1$$

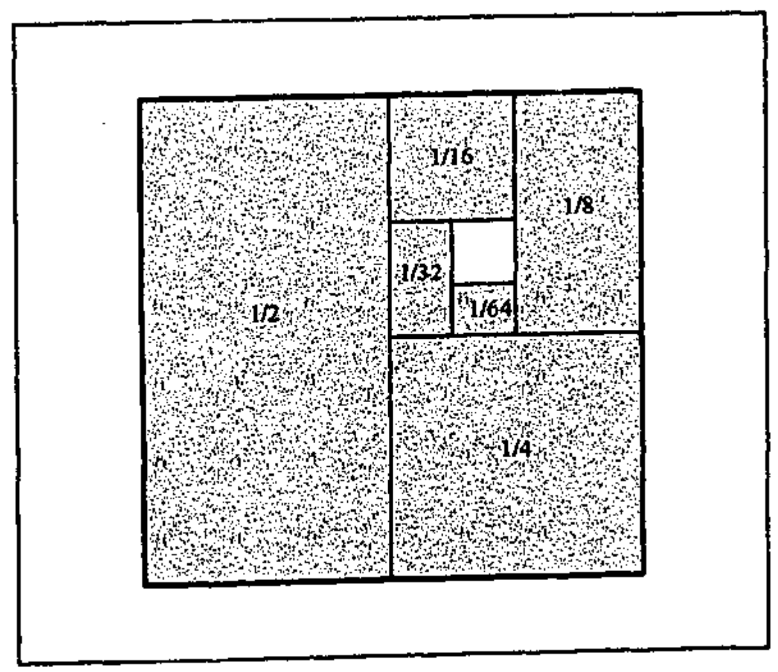


Figure 8

Of course, it is well known to mathematicians that visual reasoning, especially in analysis can be misleading. Visual intuition may also suggest that a continuous function ought to be differentiable at all but a finite number of points; the existence of continuous but nowhere differentiable functions (functions which are not fully visualizable of course) shows that here our intuition has led us astray.

But it would be a mistake to infer from this that visualization cannot provide us with any evidence for conjectures in analysis. Ordinary perception can sometimes be misleading too, but that does not prevent it from being a source of evidence. Rigorous symbolic reasoning is no doubt a more reliable source of evidence for theorems in analysis than visualization, but it is not infallible either – witness the subtle error that crept into Cauchy's attempted proof that a convergent series of continuous functions is continuous.

In 'Epistemology of Visual Thinking in Elementary Real Analysis' [1994] Marcus Giaquinto argues that in analysis, unlike the case of geometry and number theory, the unreliability of visualization prevents it from being a route to *knowledge* of analytic theorems of a certain level of generality, although he does suggest that restricted theorems can be known in this way. But even if visualization cannot provide us with knowledge of theorems, this would not seem to prevent it from at least providing a defeasible kind evidence for conjectures in analysis.

A visual argument does not rigorously demonstrate a theorem, but it can make it more plausible; the visual demonstration of the intermediate value theorem is a case in point. Furthermore, although visualization in mathematics can sometimes mislead us, as our knowledge and experience of a mathematical field improves, we gain a better understanding of when visualization is reliable and when it is not. As Giaquinto remarks:

> There is no doubt that we can be fooled by our visual intuitions. The practical moral is not that we should avoid visualizing in analysis, but that we should try to improve our grasp of the conditions under which visualization is misleading...With experience in analysis one surely develops a more discriminating attitude to the promptings of one's visual imagination; thus visualizing becomes more useful with expertise, not less.
>
> [Giaquinto 1994, p. 812]

Visualization in mathematics is not infallible and is certainly defeasible in the light of theoretical knowledge. But exactly the same point could be made with respect to observational evidence in the natural sciences. That being so, the fact that visualization can sometimes mislead us does not prevent it from being an important source of non-deductive evidence in mathematics.[108]

6. Problems

To a first approximation, the account of the development of mathematics which I have been building up to can be stated as follows. We begin with a body of mathematical results – some basic facts of geometry or arithmetic for example – which are initially accepted by appeal to the various kinds of non-deductive evidence mentioned above or on empirical or perceptual grounds. Axioms may then be introduced in order to systematize such a body of results. We saw how this account applies to the introduction of axioms for geometry, number theory and set theory. In each case we have a set of accepted results and we introduce axioms which will enable us to derive those results from a few basic principles in a systematic way. The axioms are justified by showing that they entail the

[108] See also [Brown 1997,1999, pp. 25-45] for a vigorous defence of the thesis that visual reasoning can provide us with evidence for mathematical statements and is n[illegible] simply a heuristic or pedagogical device.

results to be systematized. At the same time, some of the accepted results will be given a more rigorous justification.

We can explain this by making a distinction between a *proof* and a *derivation*. A derivation is an argument which serves to justify its *premises*, rather than its conclusion, by showing how those premises entail a result that we already accept.[109] An example would be Peano's derivation of the numerical equations and laws from his axioms for number theory. This derivation initially provided evidence for the axioms, rather than the equations and laws, which had come to be accepted on other, non-deductive grounds. As the premises of a derivation become better and better supported by this kind of evidence, the premises may come to be accepted as true. The reasoning involved in the derivation then serves to establish its conclusion, since it shows that the conclusion can be deduced from premises we now take to be firmly established. In this way, a derivation can become a proof. The original reasoning which supported the conclusion of the derivation may then be abandoned. So for example, once Peano's axioms had become accepted, the original non-deductive, quasi-empirical justification for the numerical equations and laws is replaced by the axiomatic proof.

In the same way, Zermelo's deduction of the well-ordering theorem from the axiom of choice was initially taken by many mathematicians to show only that the axiom of choice implies the well-ordering theorem. But of course, the fact that the axiom implies this result can be taken as providing a partial justification for it. That, is Zermelo's argument was initially taken to be a derivation, rather than a proof. As the power of the axiom of choice as a tool for deriving important results became clearer, the axiom gradually came to

[109] This kind of argument was recognized by Russell;

be accepted as true. Once that had happened, Zermelo's argument became a *proof* of the well-ordering theorem.[110]

In general, new first principles, axioms and definitions are adopted on the grounds that they allow us to give systematic and rigorous proofs of a prior body of accepted results. The attempt to construct such a systematic theory may reveal that some of our accepted results need to be rejected or more carefully stated. In this way, our understanding of those results is improved and more rigorous justifications for them provided. This is the process of reflective equilibrium at work in mathematics. We aim to construct our theory which accounts for as many of our beliefs as possible, although in the process, we may discover that some of those beliefs need to be abandoned.

Of course, once this has been done, the new principles, axioms and definitions may be used to deduce new results, not previously accepted. The mathematicians who came after Euclid, for example, used his axioms to establish many new geometrical theorems. In the same way, once the axiom of choice had been accepted, it could be used to prove not only the well-ordering theorem, but many new theorems in set theory. In this way, genuinely new knowledge is added to the set of accepted results.

But the process of reflective equilibrium need not end there. We will now have a new set of accepted results on which the process can start to work again. It is in this way that mathematical knowledge is continually extended and improved. Beginning perhaps,

[110] In Kitcher's framework, we can think of a derivation as a fourth kind of member of the set of accepted reasonings; an argument which is neither a proof, an unrigorous argument or a non-deductive argument, since all of these serve to establish their conclusions, rather than their premises. The derivation supports its premises by showing that they entail a statement which is itself supported by an independent argument, an accepted reasoning which is of one of the latter three kinds. There is then another kind of change which can occur in the set of accepted reasonings; a derivation may become a proof and the original reasoning which supported its conclusion is dropped from the set of accepted reasonings.

from a small collection of empirically justified mathematics, this iterative process has led to the full body of contemporary abstract mathematics.

It is a consequence of this account that we have something like a hypothetico-deductive justification of axioms in mathematics. Just as in science, we confirm a hypothesis or law by deriving empirically testable consequences from it, we confirm mathematical axioms by deriving known mathematical consequences from them.

The problem with this account, and the reason it can only be a first approximation to an adequate description of mathematical evidence, should now be readily apparent. For I argued in chapter four that the hypothetico-deductive account of confirmation is inadequate; logical entailment alone is not sufficient for confirmation. All the problems with the H-D account discussed in chapter four apply equally to the case of mathematics. The fact that a proposed axiom entails an already accepted result is not enough for that axiom to be confirmed by it. Suppose I were to propose the following new axiom for arithmetic:

(A) $$2+2=4 \rightarrow (F \& L)$$

where F stands for Fermat's Last Theorem and L for the associative law of addition. Given that we accept the antecedent, (A) entails a known result (L) and in addition we get a neat three line proof of Fermat's Last Theorem. Of course we could replace L in (A) with any number of accepted statements of number theory, including statements which we have good reason to believe are true, but which cannot be proved on the basis of the existing axioms. Our proposed axiom will then entail any number of statements for which we have a great deal of independent evidence. But surely none of this would be sufficient to justify the axiom.

Given any derivation of a known result from a set of axioms, we can add any number of arbitrary statements to the set and the derivation will still go through. Hence if the original derivation served to confirm its premises, so will the new derivation. This is just the problem of irrelevant conjunction again. Of course, we can make the same moves here that we considered before; we may require the premises of the derivation to be consistent and for them to contain no subset of statements which entail the conclusion. But we have already seen that this is not enough. Instead of (A) above, we could introduce the axiom (B):

(B) $$F \& (F \rightarrow L)$$

This axiom allows us to derive any number of known results L, it is consistent and each conjunct is indispensable in the derivation of L. As a bonus, (B) also implies Fermat's Last Theorem. But clearly such a derivation of L would not confirm (B) – in particular, it would not provide us with any evidence for Fermat's Last Theorem.[111]

We have seen that this sort of problem is a general one for the hypothetico-deductive account of confirmation. In chapter four, I argued that the solution to this problem may be to replace the H-D account with an *explanation criterion* for confirmation: if a deduction of a certain known result from a hypothesis *explains* that result, then that result counts as evidence for the hypothesis.

On this account, what is wrong with axiom (A) is not that it is not self-evident or knowable *a priori*. Rather, the problem is that it does not explain its consequences. In particular, it does not provide even a possible explanation of Fermat's Last Theorem. The same applies to axiom (B). If we take L to be the associative law of addition, it is clear that

[111] Compare this to the derivation of (U') in chapter four, section three.

the axiom does not *explain* why the law is true and that is why the derivation of L from (B) does not provide any confirmation of (B).

Axioms in mathematics are justified by showing not merely that they entail the body of results being systematized, but by showing that they provide explanations of those results. This means that we must have something like inference to the best explanation in mathematics. If this is right then a theory of mathematical explanation will be needed in giving an account of the nature of mathematical evidence.

We can arrive at the same conclusion by considering the role of non-deductive evidence in mathematics. I have argued that this kind of evidence plays a crucial role in the development of mathematics, by providing the main source of independent evidence for the statements which serve to justify axioms and other first principles. However, the summary account of the various kinds of non-deductive evidence given in section five will not quite do as it stands. We saw there that one important kind of non-deductive evidence - verification of a consequence – applies not only to the justification of axioms, but also to the justification of lower level mathematical conjectures. Clearly then, the problems described above with respect to the justification of axioms will also apply here; the fact that a conjecture entails a known result is not sufficient for it to be confirmed by that result. To the extent that induction, analogy and so on can also be thought of as special cases of verification of a consequence, my account of these forms of non-deductive evidence will face many of the same difficulties. Again, the solution may be to say that a conjecture must *explain* its consequences for it to be confirmed by them.

Consider the problems with the induction schema mentioned towards the end of section 5.1. We found that our initial characterisation of the way in which a universal

statement is confirmed by its instances had unintended consequences; that statements concerning objects which are not even numbers and not the sum of two primes (irrational numbers, equilateral triangles, coffee mugs) provide us with evidence for Goldbach's conjecture for example.

By applying Goodman's 'grue' construction, we also saw that the numerical evidence for Goldbach's conjecture equally confirms an infinite number of other, mutually inconsistent hypotheses. But notice that Goldbach's conjecture cannot be used to *explain* why certain things which are not even numbers are not the sum of two primes. We could then argue that this is why the latter observation does not confirm Goldbach's conjecture; only those statements which are explained by a hypothesis can confirm it. In the same way, we could argue that the conjecture *all even numbers greater than two are the sum of two primes* provides a better explanation of the numerical evidence than that provided by the alternative hypotheses of the form *all even numbers greater than two are* GRUE and that is why we take the numerical evidence to support the first conjecture and not the second. Clearly, in order to make these ideas more precise, we will need an account of what it means to explain a mathematical statement.

The existence of non-deductive evidence in mathematics raises an interesting further question. If we have all this evidence for mathematical statements, evidence of a kind which seems to satisfy other scientists, why do we need *proofs* at all? An answer to this question which seems promising is to say that we seek proofs in mathematics because they provide us with better *explanations* of the mathematical facts. The numerical evidence for the Riemann hypothesis for example, provides us at most with some reason for thinking *that* it is true, but a *proof* would give us an insight into *why* it is true. It is this concept of an explanatory proof which we must now investigate in the final chapter.

CHAPTER SIX

MATHEMATICAL EXPLANATION

1. EXPLANATORY PROOFS

Like other scientists, mathematicians aim to do more than simply accumulate truths. Where possible, they also seek to achieve a better understanding of their subject matter. That is, they look for *explanations* of the facts under investigation.

There is a common perception amongst mathematicians that some proofs are more explanatory than others. Some proofs of a theorem *p*, show not only *that p* is true, but also reveal to some extent, *why p* is true. That a proof provides an explanation of a theorem is considered an epistemic virtue and that a proof is non-explanatory a failing, though not one which necessarily detracts from the certainty of the theorem. This is why mathematicians often provide new proofs of old theorems. It may not be the truth of the old theorem that is in question, rather the mathematician may be attempting to give a better explanation of an already established result.

For example, the Pythagorean theorem on right-angled triangles has been proved in hundreds of different ways. E.S Loomis published a book listing 370 such proofs.[1] Only a few of these proofs however, provide anything like an explanation of why the theorem holds. We will consider one explanatory proof of the theorem in what follows.

Further examples of the distinction between explanatory and non-explanatory proofs can be given. Equations of degree 1, 2, 3 and 4 can be solved by an exact formula,

[1] See [Loomis 1968].

but the equation of degree 5, the *quintic* cannot. This fact seems to cry out for an explanation. What is so special about the number five? Abel gave the first proof of the insolubility of the quintic in 1824, but it was not until the group-theoretic proof of Galois that the result was given an explanation.[2]

Another example of the distinction, due to Mark Steiner, is provided by the prime number theorem, mentioned in the previous chapter. Recall that the first proofs of this theorem made use of sophisticated techniques of complex analysis and that an alternative 'elementary' proof was later given by Erdös. This proof is called elementary, not because it fails to be long and complicated, but because it uses only arithmetical and combinatorial notions and does not require the theory of complex analysis. Despite this, the analytic proofs are considered by many mathematicians to be more explanatory, since they provide a much better estimate for given n, of how much $\pi(n)$ differs from $n/\log(n)$.[3]

The existence of explanations in mathematics suggests a philosophical question; what is the difference between an explanatory and a non-explanatory proof? Obviously this problem is related to the more general issue of explanation in the philosophy of science. Perhaps we can look to the various accounts of scientific explanation offered by philosophers working in this field, for an answer to our question.

Many such accounts however, are clearly not going to work. Consider Hempel's *deductive-nomological* account, according to which to explain something is to deduce a sentence describing it from a set of sentences which must contain at least one general law.[4] This account of scientific explanation fails spectacularly in the mathematical case, since *all* mathematical proofs can be thought of as D-N arguments A popular type of solution to the

[2] See [Kline 1972, pp 752-771, Stewart 1987, pp. 80-92].
[3] See [Steiner 1978, pg 137].
[4] See [Hempel 1965, chapters 9, 10, 12; 1966, pp. 47-69; Salmon 1989, pp. 12-25].

difficulties which face the D-N account is to invoke causal notions. To explain something on such an account is to cite one or more of its causes.[5] Such accounts seems inappropriate for mathematics, given the assumption that mathematical objects are causally inert. Accounts of explanation which make use of modal notions or ideas from probability theory and statistics seem similarly inappropriate.

It may be that mathematical explanation is a special variety of explanation, bearing only a family resemblance to other forms of scientific explanation. Presumably however, it would be a good thing to have a unified account of scientific explanation; one which applied both to empirical science and mathematics. This suggests that philosophers of science, who want to give an account of scientific explanation, need to pay attention to examples of explanation in mathematics in addition to explanation in physics, biology and so forth.

It is perhaps surprising then, that the distinction between explanatory and non-explanatory proofs seems to have received very little attention in the philosophical literature. One of the reasons for this may be that our intuitions in this area are not very strong and so there is often disagreement on what does and what does not count as an explanatory proof. But this only makes the problem more interesting; a good account of mathematical explanation should try to explain this phenomenon.

In this chapter I will describe and evaluate various suggestions for characterizing the distinction between explanatory and non-explanatory proofs. In the next section, I consider some ideas suggested by the seventeenth century debate on the status of proofs by *reductio ad absurdum*. In section three, I look at the idea that it is the generality of a proof that serves to make the distinction.

[5] See for example [Lipton 1993].

The most fully worked out account of mathematical explanation I know of is that due to Mark Steiner. I will describe Steiner's account in section four and discuss a criticism of it made by Michael Resnik and David Kushner. Resnik and Kushner propose a purely pragmatic account of explanation in mathematics, which I examine in section five. In section six, I describe an account of scientific explanation proposed by Philip Kitcher, according to which explanation is to be characterised in terms of the concept of *unification.* I examine the prospects for using this account to make the distinction between explanatory and non-explanatory proofs. In the final section, I further develop these ideas and attempt to show how an account of mathematical explanation along these lines can be used to solve some of the problems connected with mathematical evidence discussed in the previous chapter.

2. DIRECT VS. INDIRECT PROOFS

In the seventeenth century, a debate arose amongst some mathematicians over the status of proofs by *reductio ad absurdum.*[6] In a letter to Mersenne of 1638, Descartes wrote that such proofs are '...the least esteemed and the least ingenious of all those of which use is made in mathematics'.[7] Other mathematicians criticized the Greek geometers for using proofs by *reductio*, on the grounds that such proofs do not reveal the way in which the result was initially obtained.[8]

The criticism that is most relevant to our subject however, is that of Arnauld who argued that although proofs by *reductio* are just as certain as other forms of proof, they are

[6] This debate is described in detail by Paolo Mancosu in 'On the Status of Proofs by Contradiction in the Seventeenth Century'. [Mancosu 1991].
[7] Cited in [ibid. p. 15]. See Descartes' *Oeuvres*, vol. I, p. 490.
[8] See [Mancosu 1991, pp. 23-4]

less explanatory, since although they convince us *that* a theorem is true, they do not tell us *why* it is true.[9] In the *Port-Royal Logic*, Arnauld wrote:

> The geometers are worthy of all praise in seeking to advance only what is convincing: but it would appear that they have not sufficiently observed, that it does not suffice for the establishment of a perfect knowledge of any truth to be convinced that it is true, unless beyond this, we penetrate into the reasons, derived from the nature of the thing itself, why it is true,....
>
> Those kind of demonstrations which show that a thing is such, not by its principles, but by some absurdity which would follow if it were not so, are very common in Euclid. It is clear, however, that while they may convince the mind, they do not enlighten it, which ought to be the chief result of knowledge; for our mind is not satisfied unless it knows not only that a thing is, but why it is, which cannot be learnt from a demonstration which reduces it to the impossible.
>
> [Arnauld and Nicole 1872, pp. 338,340, cited in Mancosu 1991, pp. 31-2]

Although proof by *reductio* might sometimes be the only possible means of establishing a theorem (proofs of the non-existence of certain objects in infinite domains for example) Arnauld was surely right to point out that there is often something unsatisfying about proofs by *reductio* from an explanatory point of view. This suggests the following conjecture: a direct proof is always more explanatory than a proof of the same theorem which proceeds by *reductio ad absurdum.*

This conjecture however, is false. Consider the following two proofs of the irrationality of $\sqrt{2}$. The first proof I will call the *standard proof,* although it is often referred to as the *Pythagorean proof.* Suppose that $\sqrt{2}$ is not irrational. That is, suppose we have:

(1) $\sqrt{2} = \frac{a}{b}$ - for some pair of integers a and b.

[9] See [ibid. pp. 30-3]

We can further assume that $\frac{a}{b}$ is reduced to its lowest terms, which is to say that there is no integer (>1) which divides both a and b. Squaring both sides of (1) we obtain:

(2) $\left(\frac{a}{b}\right)^2 = 2$

(3) $\frac{a^2}{b^2} = 2$

(4) $a^2 = 2b^2$

From (4) it is obvious that a^2 is even. But if a^2 is even, a itself must be even. Hence:

(5) $a = 2k$ - for some integer k.

Substituting (5) in (4), we have:

(6) $(2k)^2 = 2b^2$

(7) $4k^2 = 2b^2$

(8) $b^2 = 2k^2$

From (8) it is obvious that b^2 is even. But if b^2 is even, b must be even also. Hence:

(9) $b = 2m$ - for some integer m

But from (5) and (9) we have it that a and b are both divisible by 2, which contradicts our assumption that there is no number (>1) which divides both a and b.

Although this proof succeeds in establishing the theorem, it does not seem very explanatory. A crucial lemma required for the proof is that for all integers x, if x^2 is even, then x is also even. However, we can give a more explanatory proof of the irrationality of

$\sqrt{2}$ without the use of this lemma. Instead, we can use the so called *fundamental theorem of arithmetic*, which states that every integer has a unique prime factorization.[10]

Again, suppose that $\sqrt{2} = \frac{a}{b}$, for some pair of integers a and b. As before, this gives us:

(1) $a^2 = 2b^2$

Now consider the power of 2, as it occurs in the prime factorization of a^2 and b^2. In both, the power of 2 must be even, since $(2^n)^2 = 2^{2n}$. So we will have an even power of 2 on the left hand side of the equation. But since the power of 2 in the prime factorization of b^2 is even and the right hand side of the equation is $2b^2$, we will have an *odd* power of 2 in the prime factorization of the right hand side.

So there will be an even number of 2s on the left hand side of the equation and an odd number of 2s on the right hand side of the equation. But this means that the left hand side cannot equal the right hand side. We have a contradiction, so our initial assumption must be false; there are no integers a and b such that $a^2 = 2b^2$ and hence $\sqrt{2}$ is irrational.[11] Clearly, an exactly analogous proof will establish the irrationality of $\sqrt{p}$ where p is any prime number. A further generalization quickly yields the result that for any integer n, $\sqrt{n}$ will be irrational, unless n is a perfect square (1, 4, 9, 16, 25, 36, ...).

We have two proofs of the same theorem, of which the second seems to be more explanatory. Yet both are proofs by *reductio*. So the use of *reductio* is not by itself sufficient to make the distinction between explanatory and non-explanatory proofs, contrary to the conjecture suggested by the remarks of Arnauld.

[10] For example, 360 can be uniquely factorised as $2^3 \cdot 3^2 \cdot 5^1$. See also table 4 in chapter five, section 5.3.
[11] See also [Steiner 1978, pp. 137-8].

A closely related conjecture might be that the right way to make the distinction is to use the notion of *constructive* proof. A constructive proof of the existence of objects having a certain property must proceed by giving an example of such an object. A non-constructive proof of the same result on the other hand, might proceed by proving (perhaps by *reductio*) that not every object fails to have the property, without giving any examples. This suggests the idea that of two proofs of the same theorem, the constructive proof is the more explanatory. In a stronger form, this idea might be used to motivate intuitionism. The idea would be that only constructive arguments are explanatory and only explanatory arguments count as real proofs in mathematics.

This conjecture too, is susceptible to counter-examples. In the previous chapter, we saw how Liouville gave the first proof of the existence of transcendental numbers. Liouville begins by proving a theorem to the effect that any rational approximation of an algebraic irrational must be less accurate than a certain fixed amount. He then shows how to construct irrational numbers which can be so closely approximated by rational numbers that they cannot be algebraic. One example of such a number was mentioned in the last chapter:

$$\frac{1}{10}+\frac{1}{10^2}+\frac{1}{10^6}+\frac{1}{10^{24}}+\ldots = 0.110001000000000000000001\ldots\ldots$$

Liouville's result was followed by proofs that e and π are also both transcendental. These proofs of the existence of transcendental numbers are constructive; we infer that there are transcendental numbers by giving examples of various irrational numbers which we can prove are not algebraic.[12]

[12] See [Kline 1972, pp. 980-2] for further details of these proofs.

Compare this to Cantor's cardinality argument to the same conclusion. Recall that Cantor established that the algebraic numbers are countable. But since there are uncountably many real numbers, there must be real numbers which are not algebraic and hence transcendental. This is a highly non-constructive argument, it establishes the existence of transcendental numbers without giving a single example. Nonetheless, Cantor's proof is far more explanatory than Liouville's; it gives us a much better understanding of *why* there must be irrational numbers which are not algebraic. Indeed, as argued in the previous chapter, the explanatory power of Cantor's set-theory in this and other areas was one of the factors which led to its eventual acceptance by the mathematical community. I take this example then, to count against the idea that constructive proofs are more explanatory than non-constructive proofs. We must seek the basis of the distinction between explanatory and non-explanatory proofs elsewhere.

3. GENERALITY

An alternative account of mathematical explanation appeals to the notion of generality – the explanatory proof is the more general proof. This is an idea that Steiner considers in his paper 'Mathematical Explanation' [Steiner 1978]. He suggests several motivating examples.

Consider again the two proofs of the irrationality of $\sqrt{2}$. We can easily restate the proof which goes via the fundamental theorem of arithmetic so that it establishes the irrationality of $\sqrt{n}$ for any integer n which is not a perfect square. The irrationality of $\sqrt{2}$ is then just a special case of this more general theorem. Steiner argues [ibid. p. 138] that the same thing cannot be done with the standard proof of the irrationality of $\sqrt{2}$, because we would need to prove appropriate generalizations of the lemma required for the standard proof and these proofs become increasingly complex.[13]

Steiner also makes use of a certain proof of the Pythagorean theorem, that in any right-angled triangle, the square on the hypotenuse is equal to the sum of the squares on the other two sides. Consider an arbitrary right-angled triangle with hypotenuse c and legs a and b (figure 1).

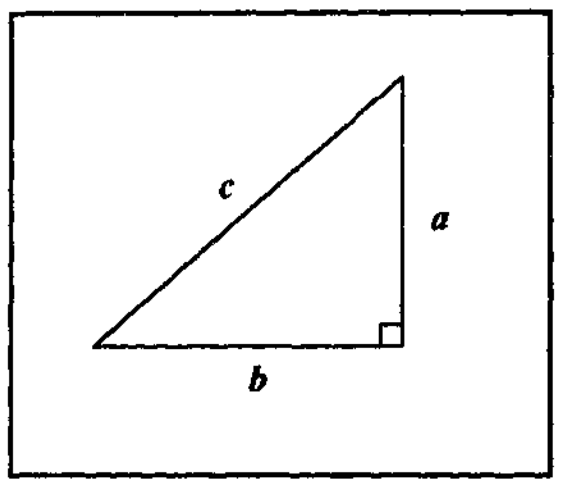


Figure 1

We want to show that $c^2 = a^2 + b^2$. What I will call the *standard proof* of this theorem makes use of the diagram shown in figure 2. This shows a square with sides of length $a + b$, which has been subdivided into four right-angled triangles congruent to the original triangle and a square whose sides are equal to its hypotenuse. Clearly the total area of the figure is equal to the sum of the areas of the triangles and the area of the centre square. The total area is $(a + b)^2$, while the area of the centre square is c^2. The area of each

[13] For example, to prove the lemma: if x^2 is even, then x is also even, we show that if we square an arbitrary odd number, the result is also odd. If $x = 2m + 1$, then $(2m + 1)^2 = (2m)^2 + 2(2m) + 1 = 4m^2 + 4m + 1 = 2(2m^2 + m) + 1$, which is clearly an odd number. We can use the same technique to show individually, for each prime number p, that if x^2 is divisible by p then so is x, but the proofs for each case get longer and longer. For the case of $p = 5$ for example, we have to square $5m + 1$, $5m + 2$, $5m + 3$ and $5m + 4$ and verify that we always get a number which is not divisible by 5.

triangle is $\frac{1}{2}ab$ (half the base times the height) and there are four such triangles. Hence we have:

(1) $(a+b)^2 = 4\left(\frac{1}{2}ab\right) + c^2$

(2) $a^2 + 2ab + b^2 = 2ab + c^2$

(3) $c^2 = a^2 + b^2$

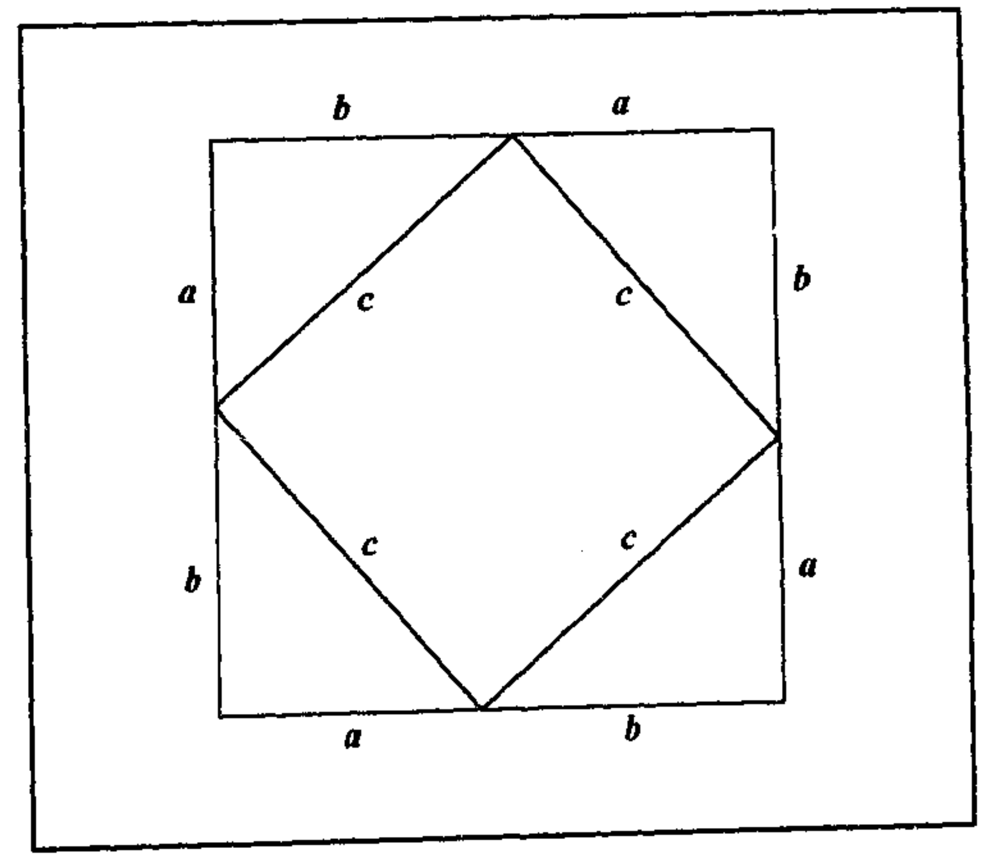


Figure 2

I hope to have convinced you that the theorem is true, but I also hope you are not satisfied with this proof as an explanation of the theorem. Steiner claims (and I agree) that the following proof is more explanatory.

There is a generalization of the Pythagorean theorem which is also true: any three similar figures[14] (not necessarily squares) constructed on the sides of a right-angled triangle are such that the area of the figure on the hypotenuse is equal to the sum of the areas on the other two sides (see figure 3).[15]

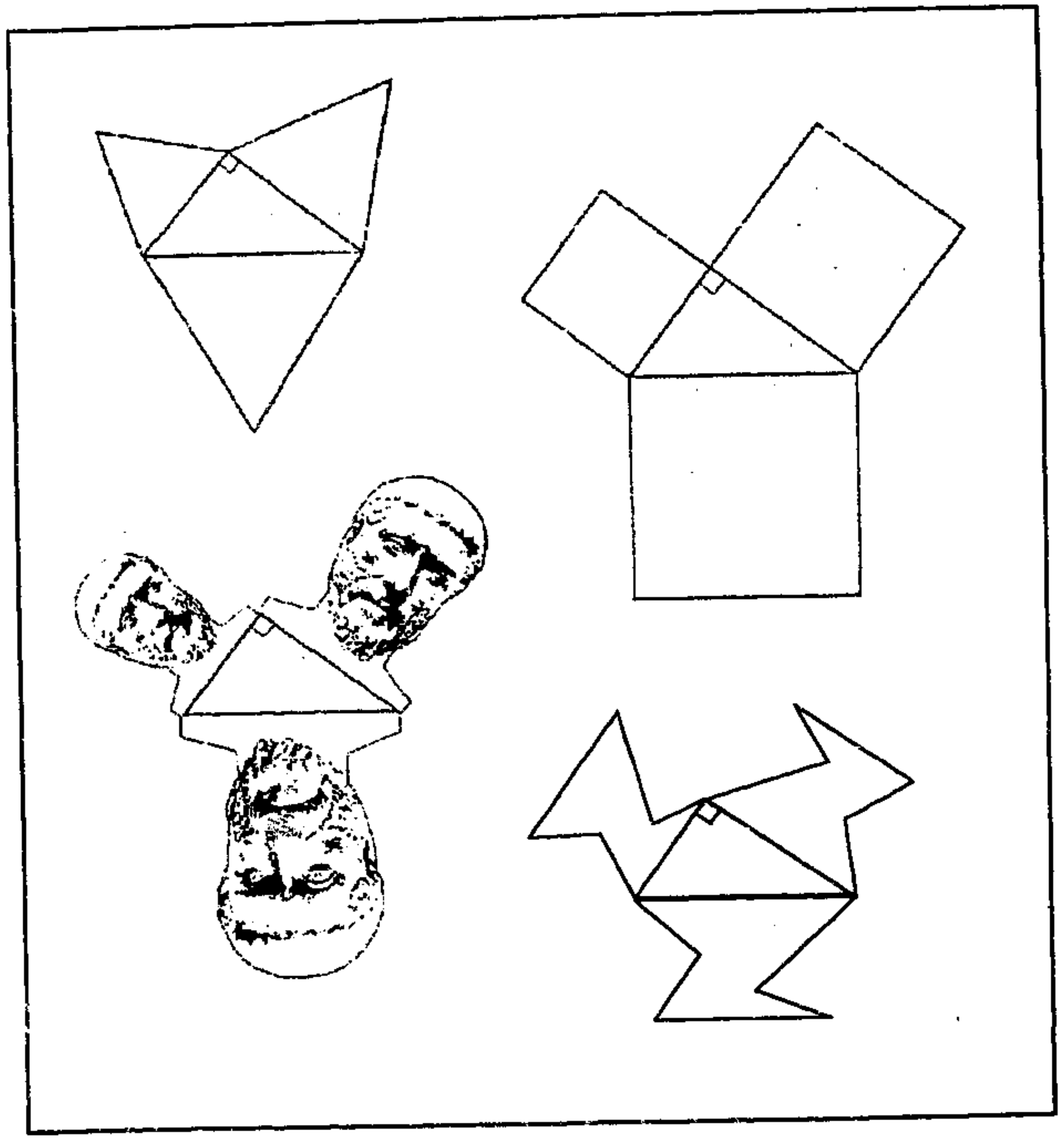


Figure 3
The Pythagorean theorem for various similar figures

[14] Recall that two figures are *similar* when the ratios of corresponding sides are equal. In other words, if we multiply all the lengths in one figure by a constant factor k, either enlarging the figure (if $k > 1$) or reducing it (if $k < 1$), then the result is a figure which is similar to the first.

[15] The right-angled triangle with the busts of Pythagoras is reproduced from [Jacob 1987, p. 412].

Clearly, if we can establish this more general theorem, the usual Pythagorean theorem will follow immediately, by taking the constructed figures to be squares.

Surprisingly though the Pythagorean theorem turns out to be equivalent to this generalization of it. Furthermore, the generalization is equivalent to *any* of its instances. Hence we can prove the general theorem (and by the equivalence this will simultaneously prove the Pythagorean theorem) if we can prove any of its instances. We do not have to restrict our attention to squares constructed on the right-angled triangle. If a proof using some other shape can be established, this will also prove the result. Indeed, one instance of the general theorem is easy to prove:

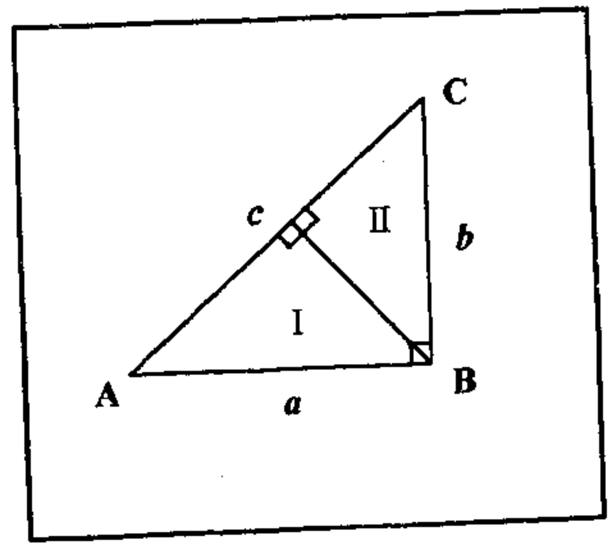


Figure 4

Figure 4 shows a right-angled triangle with an altitude drawn from the right angle to the hypotenuse. We have triangle I constructed on side a, triangle II constructed on side and the whole triangle ABC can be regarded as constructed on its own hypotenuse, c. Furthermore, it is easy to show that triangle I is similar to triangle II and that both are similar to the whole triangle ABC. So we have three similar figures constructed on the sides of a right-angled triangle and clearly the area of the figure constructed on the hypotenuse (triangle ABC) is equal to the sum of the areas of the figures constructed on the other two sides (triangles I and II). So we have established a particular instance of the general theorem, which suffices to prove the general theorem itself. The usual Pythagorean theorem then follow immediately as a special case.[16]

Steiner takes these two examples as motivation for the suggestion that '...of two proofs, the more explanatory is the more general. To deduce a theorem as an instance of a generalization, or as a corollary of a stronger theorem, is more explanatory than to deduce it directly.' [Steiner 1978, p.139]. Philip Kitcher attributes a similar claim to Bolzano, according to which '[a] proof is explanatory if and only if its premises are at least as general as its conclusion'. [Kitcher 1975, pp. 252-267]. Let us consider Bolzano's criterion first.

The suggestion is that in an explanatory proof, the premises cannot be less general than the conclusion. This suggestion seems to be well supported by the two motivating examples. In the explanatory proof of the irrationality of $\sqrt{2}$, we first prove the general result, that $\sqrt{n}$ is irrational when n is not a perfect square and then derive the irrationality of $\sqrt{2}$ as a special case. So the premises of the proof are more general than the conclusion. The same is true when we derive the Pythagorean theorem as a special case of the more general theorem.

What is less clear is that the non-explanatory proofs fail to meet the criterion. Are the premises of the standard proof of the Pythagorean theorem less general than the

[16] For more information on this proof of the Pythagorean theorem see [Polya 1954a, pp. 16-17, Steiner 1978, pp. 138-139].

conclusion? What of the premises of the standard proof of the irrationality of $\sqrt{2}$? Obviously, to answer this question properly, we need to say more precisely what we mean by one statement being more or less general than another.

An obvious answer suggests itself: A is more general than B if and only if A entails B but B does not entail A. But the obvious answer will not work. We can deduce the usual Pythagorean theorem from the general theorem and this seems to be an explanatory proof. But recall that the general theorem is *equivalent* to any of its instances. So we could also deduce the general theorem from the usual Pythagorean theorem. I take it that such a proof would be non-explanatory. However both proofs would meet Bolzano's generality criterion, interpreted in terms of deductive strength. In both cases, the premise is at least as general as the conclusion, since in fact the premise and conclusion are equivalent.[17]

The case of the generalization of the Pythagorean theorem shows that *A is more general than B* can be true even though A and B are equivalent. Perhaps some other account of generality can save Bolzano's thesis.

Kitcher argues that this is unlikely. Consider proofs by mathematical induction. Here the premises are:

(1) The number 0 has property F.

(2) For all numbers n, if n has property F, then $n + 1$ has property F.

From which we infer that:

(3) All numbers have the property F.

[17] This problem seems to be analogous to the asymmetry problem in deductive accounts of scientific explanation. See for example [Salmon 1989, pp. 46-50].

Proofs by mathematical induction occur throughout mathematics, but Bolzano's thesis seems to have the consequence that no proof by mathematical induction can be explanatory, for as Kitcher puts it, '...*whatever* account Bolzano gives ... [he] would surely be hard put to avoid the consequence that the proposition expressed by '[0] has F' is less general than that expressed by 'Every number has F'' [Kitcher 1975, p. 264].

What then of Steiner's more modest claim that 'to deduce a theorem as an instance of a generalization, or as a corollary of a stronger theorem, is *more* explanatory than to deduce it directly'? Steiner himself provides the following counter-example, which is yet another proof of the irrationality of $\sqrt{2}$. We can prove the general result that

$$a^2 = nb^2$$

implies that n is a perfect square as follows. Assume that $\frac{a}{b}$ is reduced to its lowest terms. If a prime p divides b, it must also divide b^2. Suppose then that $b^2 = kp$ for some k. Then $a^2 = nkp$ from which it is obvious that p divides a^2. But if p divides a^2, it must also divide a. Hence p divides both a and b, which is a contradiction. So no prime divides b. But the only number not divisible by a prime is 1. So b must be 1 and hence n is a perfect square (a^2 in fact). Specializing to the case where $n = 2$, we get the result that $\sqrt{2}$ is irrational as desired. But this proof, though more general, is not obviously more explanatory than the standard proof.

Steiner also considers two proofs of an identity of Euler's, concluding that the more general proof is in fact less explanatory.[18] He concludes that the generality criterion fails and this brings me to Steiner's own account of the distinction between explanatory and non-explanatory proofs.

[18] [ibid. pp 139-142]

4. STEINER'S ACCOUNT

Steiner describes his account of mathematical explanation as follows:

> My view exploits the idea that to explain the behaviour of an entity, one deduces the behaviour from the essence or nature of the entity. Now the controversial concept of an essential property of x (a property x enjoys in all possible worlds) is of no use in mathematics, given the usual assumption that all truths of mathematics are necessary. Instead of 'essence', I shall speak of 'characterizing properties', by which I mean a property unique to a given entity or structure within a *family* or domain of such entities and structures ... My proposal is that an explanatory proof makes reference to a characterizing property of an entity mentioned in the theorem, such that from the proof it is evident that the result depends on the property. It must be evident, that is, that if we substitute in the proof a different object of the same domain, the theorem collapses; more, we should be able to see as we vary the object how the theorem changes in response.
>
> [Steiner 1978, p. 143]

According to Steiner, then, a proof is explanatory if and only if it meets the following two conditions:

(1) The proof depends on a characterizing property of the objects referred to in the proof, where a characterizing property is one which uniquely picks out the object from a family of related objects. 'Depends on' here means that the proof fails to go through if we substitute another object of the same domain.

(2) By suitably 'deforming' the proof while holding the 'proof idea' constant, we can get a proof of a related result.

Let us see how Steiner's account copes with the examples given so far. Consider the explanatory proof of the irrationality of $\sqrt{2}$, which proceeds by considering the number of

2s in the prime factorization of each side of the equation $a^2 = 2b^2$. This proof makes use of a characterizing property of objects mentioned in the theorem, namely the prime factorization representation of the numbers involved. Furthermore, if we replace 2 with 4 (or any other square), the proof fails, since the prime factorization of 4, contains an even number of 2s, allowing $a^2 = 4b^2$. Finally by 'deforming' the proof in various ways, we get related theorems; the irrationality of $\sqrt{p}$ for any prime p, for example.

The explanatory proof of the Pythagorean theorem also fits Steiner's account. The proof makes use of a characterizing property of right-angled triangles, namely that they are the only triangles decomposable into two triangles similar to each other and to the whole. So substituting an acute or obtuse triangle will prevent the proof from going through. But Steiner points out that as we vary the right angle, we get related theorems, which for each such triangle tell us the difference between the sum of the squares constructed on the [illegible] and the square on the opposite side. [ibid. p. 144]. Steiner goes on to show his account applies to further examples, including the explanation of the impossibility of a sixth regular polyhedron by means of Euler's equation V – E + F = 2 and Galois's group-theoretic explanation of the insolubility of the quintic [ibid. pp. 145-50].

In 'Explanation, Independence and Realism in Mathematics' [1987] Michael Resnik and David Kushner offer several objections to Steiner's account. I will briefly describe just two of their objections. Both take the fonn of counter-examples.

The first counter-example is a proof which meets Steiner's two conditions, but is not explanatory. For this purpose, they use the standard proof of the irrationality of $\sqrt{2}$. As a characterizing property they suggest 'being the least integer x such that any integer that x divides is also divisible by 2'. This does single out 2 from the other integers and varying it

appropriately yields all the related results that Steiner notes in favour of the proof based on prime factorization [see ibid. p. 147].

Steiner criticized the standard proof on the grounds that the appropriate version of the required lemma (if x^2 is even, then x is also even) must be reproved in each case to get the related results and the proofs become increasingly complex. Resnik and Kushner remove this difficulty by proving the general lemma, that if $a = kn + i$, where $0 < i < k$ (that is, k does not divide a) and k is not a perfect square, then k does not divide a^2. [ibid. p. 147].

The second counter-example suggested by Resnik and Kushner is a proof which is explanatory, but does not meet Steiner's conditions. Here they make use of a modern proof of the intermediate value theorem, which they state in the form: if a real-valued function f is continuous on the closed interval $[a, b]$ and if $f(a) < c < f(b)$, then there is an x in $[a, b]$ such that $f(x) = c$.[19]

The proof is based two fundamental principles. The first of these is a consequence of the continuity of the real number system:

(1) Every non-empty set of real numbers which is bounded above has a least upper bound.

The second principle is simply one standard definition of what it means for a function to be continuous:

(2) A function f is *continuous at a point* x, if and only if, given any open interval J around $f(x)$, we can find an open interval I around x such that for every point y in I $f(y)$ lies

[19] In the previous chapter, we looked at the special case where $c = 0$, sometimes called the intermediate zero theorem. In fact, the intermediate zero theorem also implies the more general theorem [see for example Giaquinto 1994, pp. 809-10]. Here then, we have here another example of a generalization which is equivalent to one of its instances.

in the interval J. A function is *continuous on the interval* $[a, b]$ if and only if it is continuous at every point in the interval $[a, b]$.

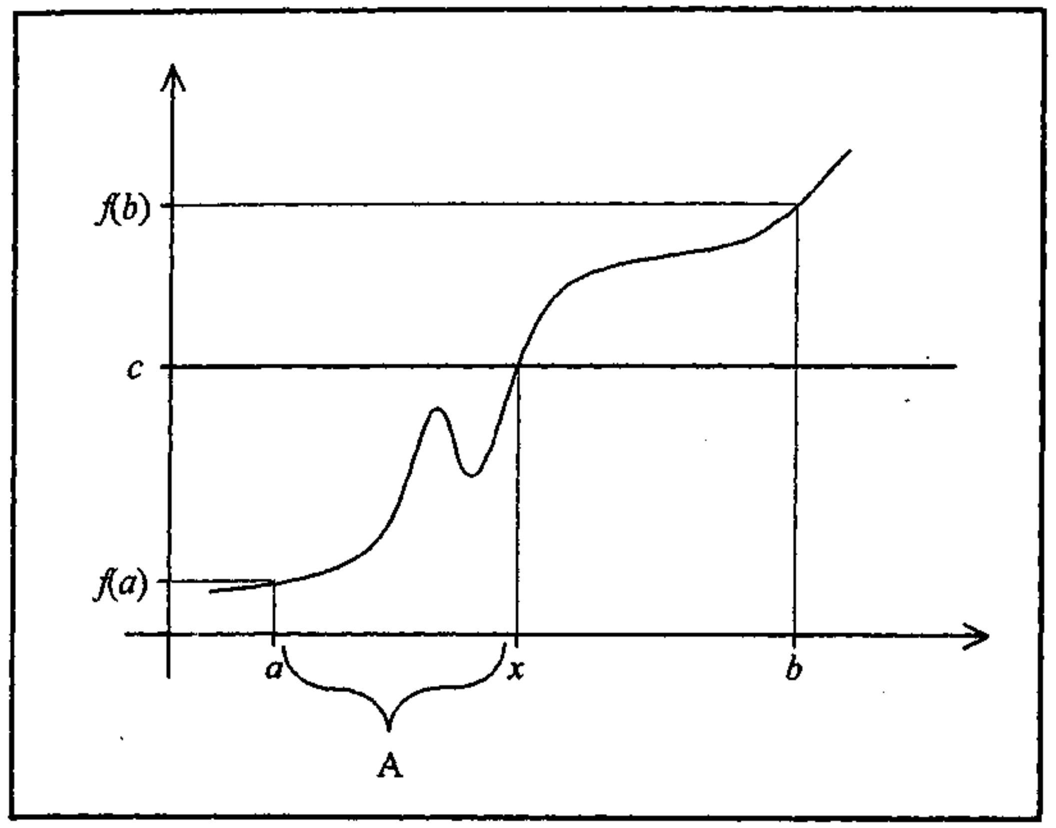


Figure 5
Proof of the intermediate value theorem

The proof is then as follows. Let f be any function continuous on the closed interval $[a, b]$ and let c be any point lying between $f(a)$ and $f(b)$ (see figure 5). We have to show that there is a point x in the interval $[a, b]$ such that $f(x) = c$. Consider the set A of all the points p in $[a, b]$ for which $f(p) < c$. The set A is not empty, since it contains a and it is bounded above by b. Hence, by (1) A has a least upper bound x. This means that every element of the set A is less than x and further, x is the *least* number with this property. Now since f is continuous on $[a, b]$, it must have a value at the point x. In fact it can be shown that $f(x) = c$.

Suppose that $f(x) < c$. Since f is continuous, (2) implies that there is an open interval I around x such that for every point p in I, $f(p) < c$. So we can pick a point y in I which is just to the right of x, for which we have $f(y) < c$ (see figure 6). But this contradicts the fact that all such points are to the left of x, since x is an *upper* bound of A.

Suppose on the other hand, that $f(x) > c$. Then again, since f is continuous (2) implies that there is open interval I around x such that for each point p in I, $f(p) > c$. Now we can pick a point z in I which is just to the *left* of x, for which we have $f(z) > c$ (see figure 6). Clearly every element of A is less than z, so z is an upper bound of A. But z is also less than x, which contradicts the fact that x is the *least* upper bound of A. Since the assumption that $f(x) < c$ and the assumption that $f(x) > c$ have both led to a contradiction, we must have $f(x) = c$.

Of this proof, Resnik and Kushner say:

> We find it hard to see how someone could understand this proof and yet ask *why* the theorem is true (or what makes it true). The proof not only demonstrates how each element of the theorem is necessary to the validity of the proof but also what role each feature of the function and the interval plays in 'making' the theorem true. Moreover, it is easy to see that the theorem fails to hold if we drop any of its conditions.
>
> [Resnik and Kushner 1987, p. 149]

However, the proof fails to meet Steiner's conditions for a proof to be explanatory:

> Although the theorem has trivial expansions and much more abstract versions … .neither the theorem nor our proof is known to be 'deformable' to yield genuinely new results. In addition, as clear as the proof is, we find it hard to identify the characterizing properties on which it depends.
>
> [ibid. p. 149]

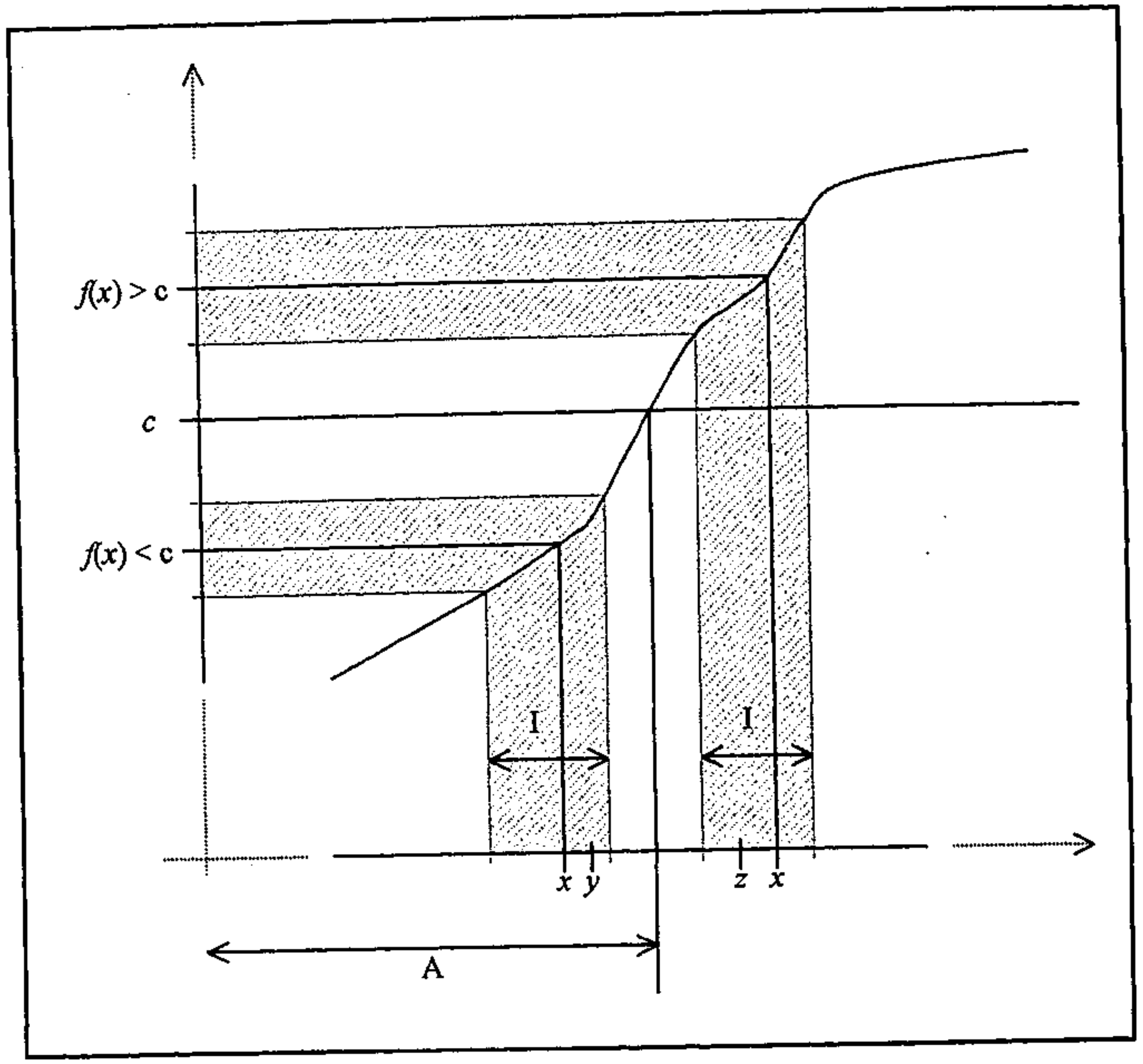


Figure 6

Of course, in a sense, the proof does depend on characterizing properties of the objects referred to in the theorem, namely the properties mentioned in the principles (1) and (2) above; (1) states a characterizing property of the real number system, which serves to distinguish it from non-continuous number systems like the rationals, while (2) states a characterizing property of continuous functions. But this is not enough to show that the proof also meets Steiner's second condition, since the proof is not generalizable or 'deformable' in the required way. That is, we do not obtain any related results by varying

the objects referred to in the proof and seeing how the theorem changes in response; if we were to replace 'real-number' with 'rational number' or 'continuous function' with 'discontinuous function', the proof would simply collapse.[20]

5. PRAGMATICS

Resnik and Kushner make some more general remarks concerning the notion of mathematical explanation. They endorse the pragmatic account of scientific explanation given by Bas van Fraassen.[21] On this account, the conditions governing the success of an attempt at explanation are entirely pragmatic. According to van Fraassen, an explanation is an answer to a *why-question*. A why-question consists of three parts; a *topic p* (where the question asks *why p*?) a *contrast-class* (why *p* rather *q*, *r*, *s* ...?) and a *relevance relation*, which determines the respect in which a given proposition would count as an answer to the question. In explaining a mathematical fact by giving a proof, the appropriate relevance relation would be the relation of logical consequence.

The context in which a why-question is asked determines its topic, contrast class and relevance relation. For example, if I ask *why is the Pythagorean theorem is true*? then in one context I might have in mind the contrast-class, *why does the theorem hold for right-angled triangles and not acute or obtuse triangles*?. Alternatively, I might have in mind the contrast class *why does the theorem hold in Euclidean space and not in various non-Euclidean spaces*? An answer to the first question may not be a good answer to the second question and vice versa. Differing contexts make different answers to the question

[20] See [ibid. p. 149]
[21] See [Van Fraassen, 1977 and 1980, chapter five]

explanatory; what is explanatory in one context may not be so in another. Resnik and Kushner put the point like this:

> ...nothing is an explanation *simpliciter* but only relative to the context-dependent why-question(s) that it answers....Whether or not a given proof counts as explanatory depends upon the why-question with which it is approached. If you simply wanted to know why a result is true (rather than false) and were prepared to accept any proof as an answer then you would count all its proofs as explanatory. But you might want to know more. For instance, in addition to wanting to know why the Pythagorean theorem holds you might want to know why it holds *only* for right triangles. Then not every proof of the theorem will contain an answer for you.
>
> [Resnik and Kushner, 1987, p. 153]

On this view, being explanatory is not an objective feature of a proof, since whether or not a proof can be used to answer a question depends on the particular why-question being asked and that depends on the interests and expectations of the questioner, for it is these which determine the topic, contrast class and relevance relation which define the question. Resnik and Kushner then account for the idea that some proofs are more illuminating than others in the following way:

> ...the so called explanatory proofs ... present more information and do so more perspicuously than do 'nonexplanatory' proofs of the same results. Thus they provide the ingredients for answering more why-questions than other proofs of the same results. But they are not explanatory in and of themselves.
>
> [ibid. p. 154]

Exactly how do we judge the relative explanatory merits of different proofs on this account? In van Fraassen's view, there are two aspects to the evaluation of explanations. Firstly, the explanation must obviously provide an *answer* to the why-question it is a response to. Given a question Q = ⟨T, C, R⟩ (where T is the topic, C the contrast class

and R the relevance relation) a *direct answer* to Q has the form 'T in contrast to C because A'. For this to actually provide an answer to the question, A and T must both be true, no member of the contrast class (other than T) must be true and A must bear the relation R to ⟨T, C⟩. In fact, asking the question Q *presupposes* that that there is a proposition A which meets these conditions. If this presupposition is false (if for example, the topic of the question is false) then the request for an explanation is rejected, for in this case there is no answer to the corresponding question.

Of course, there may be many different answers to a particular why-question, all of which meet these conditions. Furthermore, some of these answers will provide better explanations of the topic than others. Some answers to a question may be more *telling* than others. This brings us the second aspect of van Fraassen's account of the evaluation of explanations; the theory of *telling answers*. For an answer to be telling it must satisfy three conditions: (i) it must be probable in the light of our background knowledge (ii) it must make the topic more likely than the other members of the contrast class and (iii) it must be better in these respects than other potential answers.

However, David Sandborg has argued that in the case of mathematical explanation, this account is a complete failure, since it implies that any proof of a mathematical statement *p* provides a perfectly telling answer to the question *why p*? He writes:

> Consider a why-question for which the contrast class consists of mutually exclusive members, such as 'why does $1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\ldots$ converge to $\frac{\pi}{4}$, rather than some other real number?'. Let us call such a question an *exclusive-contrast question*. Under van Fraassen's theory, any proof of the topic at all will be a completely telling answer to such a question. The answer itself will be judged as maximally probable, as it follows from accepted mathematical propositions. The topic itself, having been proven, will have probability 1. All other members of the contrast class, being incompatible with the topic, will

have probability 0. No other answer can be more probable, favour the topic better, or screen it off. Therefore, van Fraassen's theory of evaluation of answers trivially recognizes any proof that establishes the truth of the topic of the question as completely telling. Thus, at least for exclusive-contrast questions, a proof must be either explanatory or not; there is no middle ground. But surely in mathematical cases, as in scientific ones, some explanations seem better than others.

[Sandborg 1998, p. 613]

This is related to a more general problem with van Fraassen's account of explanation. Notice that van Fraassen places no constraints at all on the allowable relevance relation R which can be used to answer a question. Potentially then, R can be any kind of relation we like. Kitcher and Salmon [1987] show that this completely trivializes van Fraassen's theory; given any proposition we can 'explain' it by means of any other proposition, by finding an appropriate relevance relation. Furthermore, they show how to construct the relation so that the answer is completely telling according to van Fraassen's three criteria.[22] They illustrate the point by showing how, according to van Fraassen's theory, we can explain why John F. Kennedy died on 22nd November 1963 in terms of 'astrological influence', by deducing that he died on this date from a true description of the position of the stars and planets on the date of his birth by means of an appropriate 'astrological theory'. But surely we want to say that this is not an adequate explanation, because the facts adduced in support of the fact to be explained are not *really* relevant to it. That is to say that some relations R are (objectively) more relevant than others to the explanation of certain facts. Hence van Fraassen's account cannot avoid the problem of describing an objective relation of explanatory relevance; pragmatic considerations alone are insufficient for an adequate account of explanation.

[22] See [Kitcher and Salmon 1987, Salmon 1989, pp. 135–146] for further details of the construction.

David Sandborg, in the paper mentioned above, goes on to argue that this general approach to the concept of explanation is misguided. The fundamental thesis of this approach is that an explanation is an answer to a why-question. The problem is that this entails that to explain something, we must already know *in advance* what an acceptable answer to the corresponding why-question would be; 'to understand a question is to know what would count as an answer to it.' [Belnap and Steel 1976, p. 35]. On this view, as Sandborg puts it:

...an explanation must respond to a *question*, which implies a fixed way of looking at the topic. But our initial state of puzzlement may be due to not even knowing how best to regard the topic. An explanation can gain most of its virtue by responding to this state of affairs – showing us an effective way to understand the subject-matter – rather than through any particular why-questions it happens to answer. In so far as asking a why-question fixes a way of looking at the explanandum and demands an explanation in those terms, the why-question approach will be subject to this problem.

[Sandborg 1998, p. 622].

Sandborg gives an example from the history of science: 'Isaac Newton provided an explanation for the movements of the planets, but not one which answers the question posed by his predecessors; they demanded mechanical explanations without reference to action at a distance...Indeed, it would have been impossible to specify a Newtonian answer as an appropriate answer to a question posed before the *Principia*, the pertinent concepts couldn't yet be given to indicate that kind of answer was appropriate.' [ibid. p. 622]. This suggests that an explanation may be successful even though it fails to answer any antecedently available why-questions. Rather '...an explanation may be significant because it deploys relevant conceptual resources not previously available'. [ibid. p. 622]. Hence, although *some* explanations may be best thought of as answers to why-questions, perhaps not all are.

This certainly seems to be true in the case of many mathematical explanations. When we ask why the general fifth degree equation is not solvable by means of a general formula in radicals, while the first, second, third and fourth are, we are unlikely to know in advance what an appropriate answer to this question would even look like. Galois' explanation of this fact works by showing how we can gain a new perspective on the problem by applying the concept of the *group* of permutations of solutions to the equation. From within this conceptual framework, we can give a far more illuminating account of the phenomenon than that provided by the long and complicated proof given by Abel. But who would have suspected, before Galois, that the concept of a group could be relevant to problems in the theory of equations? Certainly no one had ever framed the question 'why, in terms of the concept of a group, is there no general formula for solving the quintic?' Indeed, they could not have done so, because the relevant concepts did not yet exist. Galois' explanation is significant because it provides new conceptual tools for tackling problems, not because it provides answers to questions which could have been formulated in advance of the explanation.

We should certainly admit that there are context-dependent, pragmatic constraints on explanation. Furthermore, it may be that such pragmatic features of explanation might go some way towards explaining some of the difficulties we have in assessing mathematical proofs for explanatory relevance. Van Fraassen (and Resnik and Kushner at least in the case of mathematical explanations) claim that once we have taken into account all the pragmatic constraints on explanation, nothing substantive remains. On this view,

'explanatory power' is not an objective feature of theories and can therefore play no significant role as *evidence* for them. It is these further claims which we should reject. Despite the problems involved in characterizing an objective relation of explanatory relevance, I am not yet convinced that a more substantive account of explanation is not possible, one in which there can be context-independent features of an argument which can make it explanatory. In the next section, I consider a theory of this sort due to Philip Kitcher and assess its potential for providing an account of explanation in mathematics.

6. Explanation As Unification

In 'Explanatory Unification' [Kitcher 1981] Philip Kitcher proposes to account for scientific explanation in terms of the concept of unification. The idea that science explains phenomena by showing how they can be derived from a systematic or unified theory had of course, already been suggested by many philosophers of science, especially in discussions of theoretical explanation.[23] Hempel, for example, wrote that:

> What scientific explanation, especially theoretical explanation, aims at is not [an] intuitive and highly subjective kind of understanding, but an objective kind of insight that is achieved by a systematic unification, by exhibiting the phenomena as manifestations of a common underlying structures and process that conform to specific, testable, basic principles.
>
> [Hempel 1966, p. 83]

Kitcher's account begins from the observation that a good scientific theory explains by providing a unified account of a large and diverse range of phenomena. He cites

[23] The first explicit formulation of a theory of explanation as unification was that proposed by Michael Friedman. [Friedman 1974]. For a discussion and critique of Friedman's account see [Kitcher 1976, Salmon 1989, pp. 94-101,131].

although *some* explanations may be best thought of as answers to why-questions, perhaps not all are.

This certainly seems to be true in the case of many mathematical explanations. When we ask why the general fifth degree equation is not solvable by means of a general formula in radicals, while the first, second, third and fourth are, we are unlikely to know in advance what an appropriate answer to this question would even look like. Galois' explanation of this fact works by showing how we can gain a new perspective on the problem by applying the concept of the *group* of permutations of solutions to the equation. From within this conceptual framework, we can give a far more illuminating account of the phenomenon than that provided by the long and complicated proof given by Abel. But who would have suspected, before Galois, that the concept of a group could be relevant to problems in the theory of equations? Certainly no one had ever framed the question 'why, in terms of the concept of a group, is there no general formula for solving the quintic?' Indeed, they could not have done so, because the relevant concepts did not yet exist. Galois' explanation is significant because it provides new conceptual tools for tackling problems, not because it provides answers to questions which could have been formulated in advance of the explanation.

We should certainly admit that there are context-dependent, pragmatic constraints on explanation. Furthermore, it may be that such pragmatic features of explanation might go some way towards explaining some of the difficulties we have in assessing mathematical proofs for explanatory relevance. Van Fraassen (and Resnik and Kushner at least in the case of mathematical explanations) claim that once we have taken into account all the pragmatic constraints on explanation, nothing substantive remains. On this view,

'explanatory power' is not an objective feature of theories and can therefore play no significant role as *evidence* for them. It is these further claims which we should reject. Despite the problems involved in characterizing an objective relation of explanatory relevance, I am not yet convinced that a more substantive account of explanation is not possible, one in which there can be context-independent features of an argument which can make it explanatory. In the next section, I consider a theory of this sort due to Philip Kitcher and assess its potential for providing an account of explanation in mathematics.

6. Explanation As Unification

In 'Explanatory Unification' [Kitcher 1981] Philip Kitcher proposes to account for scientific explanation in terms of the concept of unification. The idea that science explains phenomena by showing how they can be derived from a systematic or unified theory had of course, already been suggested by many philosophers of science, especially in discussions of theoretical explanation.[23] Hempel, for example, wrote that:

> What scientific explanation, especially theoretical explanation, aims at is not [an] intuitive and highly subjective kind of understanding, but an objective kind of insight that is achieved by a systematic unification, by exhibiting the phenomena as manifestations of a common underlying structures and process that conform to specific, testable, basic principles.
>
> [Hempel 1966, p. 83]

Kitcher's account begins from the observation that a good scientific theory explains by providing a unified account of a large and diverse range of phenomena. He cites

[23] The first explicit formulation of a theory of explanation as unification was that proposed by Michael Friedman. [Friedman 1974]. For a discussion and critique of Friedman's account see [Kitcher 1976, Salmon 1989, pp. 94-101,131].

Newtonian mechanics and Darwin's theory of evolution as examples of theories which were accepted on the basis of their ability to unify and hence explain the phenomena under investigation. Newtonian mechanics allows for the motions of a wide range of objects in different situations to be calculated in essentially the same way, from a few fundamental principles, while Darwin's theory of evolution suggested that wide and varied range of biological phenomena could all be explained in terms of the operation of the process of natural selection.

On Kitcher's account, theories such as these unify our beliefs by providing a few basic *patterns* of argument which can be used to derive a large number of accepted sentences. The concept of a pattern of argument which Kitcher uses here is not quite the same as that familiar from formal logic. Arguments can be similar in at least two different respects; they may have a similar logical form (this is the sense of 'pattern' which interests the logician) or they may be similar with respect to their subject matter; that is they may share some non-logical vocabulary. The arguments which are used in unified scientific theories will generally involve both kinds of similarity; they will be represented by argument schemas which contain some non-logical vocabulary (force, mass, organism, population and so on) in addition to logical vocabulary and schematic letters.

The patterns of arguments we aim for in science are hence what Kitcher calls *stringent* patterns; arguments instantiating such are pattern are required to have a similar logical structure and to make use of a common set of non-logical concepts. So for example, Kitcher represents the pattern of argument introduced by Newtonian mechanics for deriving the equation of motion of an object α as follows:

(1) The force on α is β

(2) The acceleration of α is γ

(3) Force = mass · acceleration

(4) (Mass of α) · (γ) = β

(5) $\delta = \theta$

The final equation (5) represents the co-ordinates of α as a function of time. (1) represents the force on α as a function of its co-ordinates and of time and (2) represents the acceleration on α in terms of its co-ordinates and their time-derivates (e.g acceleration = d^2x/dt^2). Lines (1) to (4) are the premises of the argument and (5) is obtained from them by means of the techniques of the calculus. Arguments instantiating this pattern will not only have a similar logical form, they will also make use of the same theoretical concepts (non-logical vocabulary) at corresponding places.[24]

The requirement that patterns of argument should be stringent allows Kitcher to rule out *spurious unification*. We could, for example, achieve a complete systematization of all our accepted statements by introducing a theory T which employs a pattern of argument such as:

From α & T infer α

where α can be replaced by any sentence we accept. However, this pattern of argument fails to be stringent. Although arguments instantiating it will be similar with respect to their logical form, they will not all be similar with respect to their non-logical vocabulary, since any vocabulary whatever can appear in the place of α. In general, if a pattern of argument can be generalized so that we can use it to derive any sentence whatsoever, then the pattern fails to be stringent. '[a]s previous writers have insisted that genuine explanatory theories

[24] See [Kitcher 1981, pp. 515-9]

should not be able to cater to all possible evidence, I am demanding that genuinely unifying patterns should not be able to accommodate all conclusions.' [ibid. p. 529].

In general, Kitcher sees science as providing us with a set of arguments – the *explanatory store* - which can be used for the purposes of explanation. The set of arguments which can be used as explanations will obviously change over time as we modify our beliefs about the subject matter to be explained. Given that we have come to accept a certain set of statements K, the problem is to find a characterisation of the set E(K) of arguments which are acceptable as explanations of the members of K. Kitcher proposal is then that E(K) is the set of arguments which best unifies K.[25]

For any set of accepted statements K, there will be a great many possible *systematizations* of K; sets of arguments which derive members of K from other members of K by means of valid rules of inference. Given any such systematization S, there will also be many sets of argument patterns such that every member of S is an instance of some pattern in the set. From these sets of argument patterns, we choose the one which does the best job of unifying S. This set is called the *basis* of the systematization S. Kitcher then defines E(K) to be that systematization whose basis has the greatest the greatest unifying power. The unifying power of a set of argument patterns is assessed in terms of the number and stringency of the patterns in the set and the number of accepted statements in K which are generated by it. Hence the best unifier of K is that set of patterns which can be used to generate the greatest number of accepted statements in K from the fewest, most stringent patterns. Then the set E(K) of arguments which are acceptable as explanations of members

[25] [ibid. p. 512]

of K will be the systematization whose basis is the best unifier of K. A particular argument will count as an explanation if and only if it is a member of E(K).[26]

Kitcher's account has several attractive features. For example, he shows how it avoids many of the problems which beset the deductive-nomological account of scientific explanation. Consider the problem of asymmetry. We can deduce the period of a simple pendulum from a specification of the length of the pendulum and the mathematical law relating the two. We can also deduce the length of the pendulum from the period in exactly the same way. However, only the latter argument would explain its conclusion. Kitcher argues that his account can explain this asymmetry:

> We have general ways of explaining why bodies have the dimensions they do. Our practice is to describe the circumstances leading to the formation of the object in question and then to show how it has since been modified. Let us call explanations of this kind "origin and development derivations"... Suppose now that we admit as explanatory a derivation of the length of a simple pendulum from a specification of the period. Then we shall either have to explain the lengths of *non*swinging bodies by employing quite a different style of explanation (an origin and development derivation) or we shall have to forego explaining the lengths of such bodies ... Admitting the argument which is intuitively nonexplanatory saddles us with a set of arguments which is less good at unifying our beliefs than the set we normally choose for explanatory purposes.
>
> [ibid. p. 525]

Kitcher also shows how his theory can deal with the problem of *irrelevance*, which the D-N account is susceptible to and how it rules out 'explanations' which proceed from accidental generalizations.[27] For my purposes, the account is attractive because it makes no

[26] [ibid. pp. 519-22].

[27] See [ibid. pp. 522-6]. The problem of irrelevance can be stated as follows. Given any general law of the form $\forall x(Fx \rightarrow Gx)$ – 'all metals expand when heated' for example – we can introduce a new generalization of the form $\forall x\ ((Fx\&Hx) \rightarrow Gx)$ – 'all metals which have had a spell cast on them by Merlin the magician expand when heated'. We can then 'explain' why this iron bar expanded by deducing it from the fact that the bar is a metal, that it was heated and that Merlin has cast a spell on it. Although such a deduction would fit the D-N account, it fails to be an explanation because the premise 'Merlin cast a spell on the iron bar' is really

use of the concept of causation or related notions. This opens up the possibility that the account will work for mathematical as well as empirical explanations and hence promises a unified account of scientific explanation.

Let us see then, how well Kitcher's account does with some of the examples of mathematical explanation we have been considering. The idea would be that a proof is explanatory if and only if it instantiates a pattern of argument which could be part of that set of patterns which provides the best unification of the relevant branch of mathematics.

Now the pattern instantiated by the proof of the irrationality of $\surd 2$ via the fundamental theorem of arithmetic can be seen as a better unifier than that of the standard proof, since exactly the same pattern of argument can be used to generate the same conclusion for any prime number and indeed for any other integer which is not a perfect square. As for the proof of the Pythagorean theorem, this too can be seen as instantiating a unifying pattern, since the same pattern (where we deduce the Pythagorean theorem from the general theorem) can be used to prove analogous results for similar figures other than squares.

Furthermore, the explanatory proof of the intermediate value theorem given by Resnik and Kushner seems to fit Kitcher's account, for the argument instantiates a pattern common to many used in a systematic and well-unified theory, namely analysis. Similar remarks apply to Cantor's proof of the existence of transcendental numbers and the group-theoretic proof of the insolubility of the quintic. Analysis, set-theory and group theory are all examples of theories which unify mathematics, providing patterns of argument which can be used to derive many and diverse mathematical statements.

---

irrelevant to the effect to be explained. Notice that this is analogous to the problem of irrelevant conjunction for the H-D account of confirmation, discussed in chapter four.

Consider Cantor's proof. He shows that there must be real numbers which are not algebraic by establishing that while there are uncountably many real numbers, there are only countable many algebraic numbers. This general *pattern* of proof turned out to be very widely applicable. To show that there are Fs which are not Gs, show that there are uncountably many Fs but only countably many Gs. So for example, we can show that there are properties of natural numbers which cannot be determined by means of a mechanical decision procedure as follows. Firstly, there must be uncountably many properties of natural numbers, since any such property can be represented by a function from the natural numbers to the set {True, False} and the set of all such functions has the cardinal number $2^{\aleph_0} = c$. However there can be only countably many *decidable* properties of natural numbers, since the set of all possible decision procedures can be enumerated. Hence not every property of the natural numbers is decidable.

I have here only very roughly sketched how Kitcher's account would apply to mathematics. It is clear that more work would need to be done for it to become convincing that the account will succeed. In particular, we would need to show how explanatory theories in mathematics have the features Kitcher describes. That is, we would have to show how such theories allow for the derivation of a large number of accepted statements by means a few stringent patterns of argument. We would need to exhibit those patterns in more detail and show how the proofs which are intuitively classified as explanatory can be seen as instantiating such patterns. I gave one illustration of such a pattern above; the pattern of cardinality arguments provided by Cantor's set theory. The discussion of the development of geometry, analysis and number theory in the previous chapter should make it plausible that these theories can also be seen as unifying mathematics by providing patterns of argument which can be used to derive a wide range of accepted statements.

Some additional support for this account of mathematical explanation can be found by looking at another set of examples. Recall the distinction discussed in chapter three, between *problem-solutions* and *confirmation techniques.*[28] A problem-solution is a method that allows us to discover answers to questions, while a confirmation technique only allows us to confirm that a given answer is correct. The technique of proof by induction is an example of a confirmation technique. If we have a question like 'what is the sum of the first n natural numbers?' mathematical induction will not generate an answer for us, although as we saw in the previous chapter, we can use it to confirm that an answer obtained by some other means is correct. There is in this case, a problem-solution technique that will generate the answer to this question for us. It can be shown that the sum to n terms of an arithmetical progression with first term a and common difference d, is given by the equation, $S = \frac{n}{2}[2a + d(n-1)]$. To answer our question, we simply put $a = d = 1$, and rearrange to get $S = \frac{1}{2}n(n+1)$. This technique thus generates an answer to our question, even if we start from a condition of ignorance, unlike the technique of mathematical induction.

Notice that we can use the same technique to generate answers to further questions of the same form. If we ask, for example, 'what is the sum of the first n odd numbers?', we can find an answer by setting $a = 1$ and $d = 2$ in the above equation to get $S = \frac{n}{2}[2 + 2(n-1)]$ and then rearranging to get $S = n(1 + (n-1)) = n^2$. Hence our method

[28] See [Kitcher 1984, pp. 182-3].

tells us that $1+3+5+7+\ldots+(2n-1) = n^2$ a result which can be confirmed, but not discovered by, a proof using mathematical induction.

In the same way, if we have already conjectured that the roots of the equation $x^2 + 7x + 12 = 0$ are –3 and –4, we can confirm this answer by substitution: we can check that $(-3)^2 + 7(-3) + 12$ and $(-4)^2 + 7(-4) + 12$ are both equal to zero. But the technique of substitution will not give you the answer if you have not already guessed it. A problem-solution technique for this sort of problem would be to use the quadratic equation $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ to generate the answer by setting $a = 1$, $b = 7$ and $c = 12$.

What is significant about this distinction is that in general a problem-solution, seems to provide an explanation of a result, while a confirmation technique does not. Consider the technique of substitution for checking solutions to equations. Clearly there is no problem here about the validity of the reasoning, but one might say that substitution is not a *proof* but a *verification* (or demonstration) that the equation has certain roots, since it does not tell us *why* the equation has those roots, in the way that the argument which makes use of the quadratic equation does. Similarly, the proof by induction confirms *that* the sum of the first n natural numbers is ½n(n+1) but the proof from the general formula for the sum of any arithmetic progression give us some insight into *why* this is true.

Since problem-solutions and confirmation-techniques need not be rigorous arguments, this distinction will also allow us to account for explanatory arguments which are not proofs. Euler's analogical technique for finding the sums of infinite series (a problem-solution technique) is certainly more explanatory than the verification by computing partial sums (a confirmation-technique) even though neither are proofs.

Computing partial sums provides us with sum evidence *that* the sum is correct, but the analogical argument goes some way toward explaining it.

The connection between this distinction and the idea of explanation as unification can be made in the following way. As we have seen, a problem-solution will instantiate a general *pattern* of argument which can be used to derive solutions to a variety of similar problems. We can use the problem-solution technique offered by the arithmetic progression formula for example, to derive results concerning the sum of the first n odd numbers, as well as many other similar results. In the same way, the problem-solution technique provided by quadratic equation provides us with a way of finding the roots of any equation of this form. A problem-solution therefore corresponds to a pattern of argument which allows to answer many related questions in a unified way. Hence, arguments which make use of a problem-solution technique are more explanatory than arguments which make use of a confirmation-technique because they instantiate a pattern which is a better unifier of the set of accepted mathematical statements.

We can also appeal to this distinction to explain why there is a *prima facie* case for saying that proofs by *reductio* are less explanatory than direct proofs. For notice that the general pattern of proof by *reductio* is a confirmation-technique, rather than a problem-solution. To describe the general pattern of proof by *reductio*, we would say something like: 'assume the negation of the statement to be proved, then attempt to derive a contradiction'. But clearly, in order to do this, we have to know what statement to negate. That is, we must already have an answer to our question to hand. A proof by *reductio* will then confirm that our answer is correct, but will not generate it for us, in the way that a direct proof might.

On the other hand, as we have already seen, this does not mean that no proof by *reductio* can be explanatory. We can explain this as follows. Whether or not an argument provides a problem-solution depends on the particular question we are asking. If the question is just 'is p true?' then *reductio is* a problem-solution – for in many cases (not all of course) we can use it to answer this question, even if we do not already know the truth-value of p. But if the question is more specific, then *reductio* is not a problem-solution. If our question has the form 'what is the value of $\phi(a)$?' where $\phi(x)$ is some functional expression, then unless we have already conjectured a value for $\phi(a)$, that $\phi(a)=b$ say, we will not be able to begin the proof by *reductio* by assuming that $\phi(a) \neq b$. In general, problem-solutions provide better explanations than confirmation techniques because they allow us to answer more questions.[29]

Similar remarks apply to various non-deductive patterns of argument. Consider the patterns of verification of a consequence and inductive confirmation of a universal statement. Clearly, we need to already have the statement to be confirmed to hand, before we can derive any consequences from it or verify any of its instances. Hence, for many questions, these patterns will not provide a problem-solutions but only a confirmation-technique. So we can say that in general, an argument which instantiates a problem-solution pattern will be more explanatory than an argument by induction or verification of a consequence.

However, the distinction between problem-solutions and confirmation-techniques is not enough to account for all of our judgements concerning the relative explanatory merits of different proofs of the ne theorem. Even among proofs which are not problem-

[29] Another point to notice is that since a problem-solution allows us to generate related results using the same pattern of argument, we can explain why the idea behind Steiner's account – that an explanatory proof must be *generalizable* – has the appeal that it does.

solutions (with respect to a certain question) we can still distinguish between better and worse explanations, as we saw in the case of the two proofs of the irrationality of $\sqrt{2}$. On the other hand, given two arguments, both of which instantiate a problem-solution pattern, one may be more explanatory than another.

I would suggest that the source of our difficulties in accounting for all the many examples of explanation in mathematics is the attempt to find a hard and fast distinction between explanatory and non-explanatory proofs or arguments. Notice first of all that any proof seems to be explanatory relative to an inductive argument to the same conclusion. The inductive evidence for the prime number theorem for example, gives us some reason for thinking that it is true, but any of the known proofs of the theorem provide us with a better explanation of why it is true. Showing how a statement can be deduced from axioms or first principles contributes to the unification of our mathematical beliefs by showing how the statement can derived from a well-established, systematic theory, which allows for the derivation of large number of diverse theorems from a few basic principles. In this sense, any proof of a theorem will contribute more to the unification of mathematics and hence be more explanatory than an inductive argument for the same conclusion.

But not all proofs need be from axioms for them to improve out understanding and provide explanations – Cantor's proof of the existence of transcendental numbers is a good example. Even within an axiom system, we can distinguish between better and worse explanations of the same theorem. Both proofs of the Pythagorean theorem mentioned in section three can be derived in Euclid's axiomatization of geometry, but one is more explanatory than the other.

The concept of an explanatory proof is highly context dependent. No proof can be said to explanatory or non-explanatory in isolation. A proof is only explanatory or not relative to another proof of the same theorem. Polya's proof of the Pythagorean theorem is more explanatory than the standard proof, but the standard proof is certainly more explanatory than many others which have been given and Polya's proof might appear non-explanatory compared to others.

The foregoing remarks suggest that in seeking an account of mathematical explanation, we should not concentrate exclusively on proofs, but should consider arguments in general, non-deductive and unrigorous arguments as well as proofs. Furthermore, rather than attempting to classify mathematical arguments as either explanatory or non-explanatory, we should first try for something less ambitious; given two arguments A and B for the same conclusion *p*, we should seek to define the comparative relation 'A gives a better explanation of *p* than B', or more briefly, 'A is more explanatory than B'.[30] Following up on Kitcher's lead, I suggest that in general we have:

A is more explanatory than B if and only if A instantiates a pattern of argument which is a better unifier of the relevant set of accepted statements than B.

We can then make further distinctions concerning the different ways in which a pattern of argument can contribute to the goal of unification. One way in which a pattern of argument can be a better unifier than another is by providing a problem-solution technique as opposed to a confirmation-technique. Hence an argument which derives its conclusion

[30] Given such a comparative relation, we could then define the concept of explanatory proof by saying that a proof is explanatory *simpliciter* iff it is the *best* explanation of its theorem, that is, if its more explanatory than any other proof of the same theorem. This would of course make all proofs of theorems which have only one proof explanatory, for such proofs would trivially be the 'best explanation' we have. But this is perhaps not too serious a problem; when we say of such a proof that it is not explanatory we mean something like 'there could be a more explanatory proof' – and we start looking for one.

by means of a problem-solution technique will generally be more explanatory than an argument which derives its conclusion using a confirmation-technique.

On the other hand, an argument which shows how its conclusion can be deduced from established principles (a proof) will clearly do more for unification than an unrigorous or inductive argument for the same conclusion, since it will provide us with a clearer view of the logical relationships which hold between members of the set of accepted statements. Of course, the unification provided in this way may be more or less systematic, depending on the state of development of the theory in question. The more systematic and rigorous the theory, the better unification we will have. So even though a proof will generally be a better explanation, in this respect, than a non-deductive argument, some proofs will still be better than others. A proof in a mature, rigorous, axiomatic theory for example, will be more explanatory in this respect than a proof in a developing theory, in which assumptions are not explicitly stated and some of the arguments are unrigorous.

Notice that these two ways in which one argument may be more explanatory than another are independent. One argument may be unrigorous and yet provide a problem-solution technique. Another argument to the same conclusion might be rigorous, although it provides only a confirmation-technique. So in one respect the first argument will be less explanatory than the second (on the rigorous/unrigorous dimension) while in another respect (the problem-solution/confirmation-technique dimension) it will be more explanatory than the second. In assessing the relative explanatory merits of two such arguments, we need to weigh up the degree of unification contributed in each of these respects in order to find the overall contribution to unification achieved by each proof. It may happen that the contribution to unification achieved through the provision of a

problem-solution technique outweighs the contribution to unification achieved by the provision of a rigorous argument. In that case, the first argument would be more explanatory overall than the second. Of course, it may turn out the other way; the unification achieved by rigor might be greater than the unification achieved by the provision of a problem-solution for a set of questions. Then the second argument would be more explanatory overall than the first. Then again, it might be that overall, the two arguments come out roughly equal. Then we cannot say that either argument is more explanatory than another, although we can still say that the first argument is more explanatory than the second in one respect, while the second is more explanatory than the first in a different respect.

We have seen then that arguments may contribute to unification and hence be explanatory in a number of different ways. An argument may be more explanatory than another in some respects and not in others.[31] Other things being equal however, a proof will be a more explanatory than a non-deductive argument and a proof in a systematic, axiomatized theory will be more explanatory than a proof in a developing, unaxiomatized theory. Among proofs in the same axiomatic system we can distinguish between those that provide problem-solutions and those that provide confirmation-techniques for sets of questions. We can make yet finer distinctions between proofs which instantiate a member of a set of patterns which provide us with varying degrees of unification, along the lines suggested by Kitcher's account. Recall that the unification achieved by a set of patterns is measured in terms of the number of patterns in the set (the fewer the better) the stringency

[31] I would argue that this is in fact an advantage of my account, for it explains why there is often some confusion in our intuitive judgements about what counts as an example of an explanatory proof. People disagree about the explanatory value of different proofs because they are focusing on differing respects in which one proof may be more explanatory than another.

of the patterns (the more stringent the better) and the number of accepted statements we can derive as conclusions of arguments instantiating some pattern in the set (the more accepted statements we can account for the better). Obviously one proof can do better than another in any other these three respects *independently*. Again we assess different proofs for relative explanatory power by weighing up all these different factors, to arrive at an overall contribution to unification. Having taken all the factors into consideration, we may find that one proof may then contribute more to unification overall than another. In that case, the first proof will be more explanatory than the second.

We have already seen how this account does quite a good job of account for the examples discussed in previous sections. However, an account of mathematical explanation ought to do more than simply allow us to account for the difference between explanatory and non-explanatory proofs. As I argued at the end of the last chapter, we require an account of mathematical explanation in order to develop an adequate account of evidence in mathematics. Let us now see how well our account of explanation as unification does on this score.

7. Explanation and Evidence

Notice first of all that we are now in a position to give a more detailed answer to the question raised at the end of the last chapter, concerning the *point* of proof. Mathematicians prefer proof to mere inductive confirmation of a conjecture because a proof provides a better explanation of a statement than non-deductive confirmation of it. Proofs achieve this, on my account, by contributing to the unification of the subject matter under investigation. This is not to deny that mathematicians also prove theorems because by doing so they can

be more sure that the statement is true. By deriving a statement from well established theory, we achieve a greater degree of certainty. But this is no absolute certainty, for the axioms on which the proof is based will themselves have been established on the basis of non-deductive evidence.

As we saw in the previous chapter, the evidence for the first principles of mathematical theories is that they can be used to derive statements which can be independently verified. However, as we have seen, that a statement *entails* a known result is not enough for it to be confirmed by it; the simple hypothetico-deductive account of this kind of evidence is inadequate. I have suggested an alternative to this account, the thesis that it is what is *explained* by a statement that provides the evidence for it. Hypotheses are confirmed not simply because they entail statements which can be independently verified, but because they provide us with explanations of them. Now the thesis that explanation is unification fits quite neatly into this account of evidence. Certainly the derivations of known results from axioms and first principles described in the pervious chapter will explain those results on this account of explanation. By showing how our accepted results can be derived from a few basic principles in a systematic way, we gain an improvement in the unification of those results and thereby provide explanations of them. Hence the verification of those results confirms the principles used to derive them.

Consider the case of De Moivre's theorem. I argued that this provided evidence for the theory of complex numbers, by allowing for the systematic derivation of a large number of independently verifiable results. Notice that here we use the same basic *pattern* of argument (substitution in De Moivre's theorem of a particular value of n, multiplying out one side of the equation, then comparing coefficients) to derive the formulae for the sine and cosine of a given multiple of an angle. Hence, De Moivre's theorem provides us with a good *explanation* of these properties of the trigonometric functions. So our account supports the judgement that in this case, the derivation of a known result from a hypothesis provides us with evidence for it.

What is less clear is that our account can avoid the problems of spurious confirmation which beset the hypothetico-deductive account. Can our account distinguish between those derivations which confirm their premise and those that do not? To answer this question, we need to show that the derivations which intuitively fail to confirm their premises are not explanatory according to our account of explanation.

Consider the problem of irrelevant conjunction. We can deduce a known result from a set of premises which includes many irrelevant statements. Intuitively, the result is not an explanation.[32] Can our theory account for this?

Here I think we can use the same general strategy that Kitcher employs in showing that his account of explanation avoids the problems of asymmetry, irrelevance and accidental generalization. Given an argument that we want to show is not explanatory, we try to show that any set of argument patterns instantiating it could not provide us with the best unification of our beliefs according to the criteria of the fewest number of most stringent patterns yielding the greatest number of accepted statements.[33]

Suppose then that we have a set of axioms $\alpha_1, \alpha_2, \alpha_3 \dots \alpha_n$ which allow us to deduce the known result β. We then add some irrelevant statement γ to our set of axioms, to

[32] This is a general problem for accounts of scientific explanation of course. On the D-N account, given any deduction of a statement from a set of premises which includes one law, we can add as many irrelevant premises as we like and the conclusion will still follow. But the deduction will then fail to be an explanation. See [Salmon 1989, p. 102].

[33] See [Kitcher 1981, p. 523].

obtain another equally valid derivation of the same result. So we have two arguments to the same conclusion:

(1) $\alpha_1 \,\&\, \alpha_2 \,\&\, \alpha_3 \,\&\, \dots\, \alpha_n \vdash \beta$

(2) $\alpha_1 \,\&\, \alpha_2 \,\&\, \alpha_3 \,\&\, \dots\, \alpha_n \,\&\, \gamma \vdash \beta$

Suppose we were to admit the argument in (2) as an explanation of β. How we would we then explain the other known results that we can also derive from our axioms? For definiteness, suppose that $\alpha_1, \alpha_2, \alpha_3 \dots \alpha_n$ are the Dedekind-Peano axioms for number theory and that β is the statement $2 + 2 = 4$. In the first argument, we derive this result from the axioms in the usual way. The second argument is exactly like the first except that it contains some redundant premise γ. The question is then, if we were to use this second argument to derive $2 + 2 = 4$, what kind of argument would we use to derive other numerical equations, such as $7 + 5 = 12$? If we use the usual derivation from the axioms without the additional irrelevant premise γ in this case, then we will have two different patterns of argument instead of one for deriving the same set of conclusions. So the set of arguments we admit as explanatory will fail to provide us with the best unification of our beliefs.

On the other hand, if we adopt argument (2) as a general strategy, then although we will have one pattern of argument that can be used to derive the same number of conclusion as the first, the pattern will fail to be stringent. For in principle, we can replace γ with any statement whatsoever without affecting the ability of our pattern to derive statements we accept. So in either case, adopting argument (2) will fail to provide us with the best unification of our beliefs and will therefore fail as an explanation.

What of the non-explanatory axioms mentioned at the end of the last chapter? These were:

(A) $2+2=4 \rightarrow (F \,\&\, L)$

(B) $F \,\&\, (F \rightarrow L)$

Axiom (A) is meant to be confirmed because it entails the known result L – the associative law of addition. But clearly adding (A) to our current set of axioms for number theory would result in a less unified system of beliefs, for we can already derive L from the existing set of axioms. Our current axiom system provides us with a general pattern for deriving results such as (L) – we use the recursive definitions of the operations concerned and apply the principle of mathematical induction. If we were to admit the argument corresponding to (A) as a way of deriving the particular law L, then if we explain other laws in the usual way, we will be using an inflated set of patterns of argument; one pattern for the case of L and another pattern for other laws.

If on the other hand, we propose to explain all laws L in this way, the pattern involved will be:

(1)	$2+2=4 \rightarrow (F \,\&\, \alpha)$	Axiom
(2)	$2+2=4$	Theorem
(3)	$F \,\&\, \alpha$	From (1) and (2)
(4)	α	From (3)

where α is to be replaced with the law we want to explain. But as before, such a pattern will fail to be stringent, since it can be generalized so as to allow for the derivation of any sentence whatsoever in the place of α. Hence the unification achieved by this pattern of argument is spurious and the derivation fails to be an explanation.

Similar remarks apply to (B). If we use pattern B to explain the associative law for addition and use the standard pattern to explain other laws, we will have an inflated set of patterns. So introducing axiom (B) will result in a less unified set of arguments for deriving accepted statements. But if we propose to explain all laws on the pattern:

(1)	$F \,\&\, (F \rightarrow \alpha)$	Axiom
(2)	F	From (1)
(3)	$(F \rightarrow \alpha)$	From (1)
(4)	α	From (2) and (3)

where α can be replaced by any law, then pattern will not be stringent, since it can be generalized to allow for the derivation of any statement whatsoever. A pattern of argument which in principle, would allow us to derive any conclusion at all achieves only a spurious unification of our beliefs. Although arguments instantiating this pattern share a common logical form, the non-logical vocabulary here is idling; it plays no significant role at all in the derivation of the conclusion from the premises, since essentially the same pattern can be used to derive any conclusion.[34]

We can use the same idea to give a better account of inductive evidence in mathematics. The simple account of induction implies that any statement of the form $\sim Fa \,\&\, \sim Ga$ confirms $\forall x(\sim Gx \rightarrow \sim Fx)$ and hence confirms $\forall x(Fx \rightarrow Gx)$. So the fact that $\sqrt{2}$ is not an even number greater than two and not the sum of two primes provides us with evidence for Goldbach's conjecture. But then in the same way, *any* object which is neither an even number greater than two nor the sum of two primes will also provide us with inductive confirmation of Goldbach's conjecture.

[34] See also [Kitcher 1981, p. 528]

We can think of this problem in the following way. From Goldbach's conjecture and the auxiliary statement that 6 is an even number greater than two, we can infer that 6 is the sum of two primes. Verification of this latter statement then provides us with some inductive support for Goldbach's conjecture. But from Goldbach's conjecture and the auxiliary statement that $\sqrt{2}$ is not the sum of two primes, we can infer that $\sqrt{2}$ is not an even number greater than two. Intuitively, this fails to confirm Goldbach's conjecture. I propose to account for the difference by saying that the first derivation provides a better explanation of its conclusion than the second[35]. Does our account of explanation as unification support the proposal?

Consider first, the argument:

All even numbers greater than 2 are the sum of two primes

6 is an even number greater than 2

Therefore: 6 is the sum of two primes

Now clearly this is not a very *informative* explanation of why 6 is the sum of two primes. A proof of Goldbach's conjecture would improve it, because it would presumably give us some insight into the connection between *being even* and *being the sum of two primes*. But compare this argument to:

All even numbers greater than 2 are the sum of two primes

$\sqrt{2}$ is not the sum of two primes

Therefore: $\sqrt{2}$ is not an even number greater than 2

It is clear that this argument does not provide us with an *explanation* of its conclusion. If we were to ask why $\sqrt{2}$ is not an even number, the obvious thing to say would be that $\sqrt{2}$ is

[35] I recently discovered a very similar solution to the paradox of the ravens in Alexander Bird's *Philosophy of Science*. Bird also makes use of the concept of unification in his account of confirmation as explanation. See [Bird 1998, pp. 91-4].

not an even number because there is no positive integer x such that $\sqrt{2} = 2 \cdot x$. Let us say that this is the standard pattern of argument we use in such cases. If we now admit the above argument as an alternative explanation for this same conclusion, then how we will explain why $\sqrt{3}$ for example is not an even number? If we use the standard pattern in this case, then we will be using an inflated set of patterns of argument and so this set will not be the best unifier of our beliefs. But if we propose to use the general pattern:

All even numbers greater than 2 are the sum of two primes

α is not the sum of two primes

Therefore: α is not an even number greater than 2

then we will be using a less stringent pattern than the standard pattern, since here α can be replaced by an expression referring to any object at all, provided only that it is not even and not the sum of two primes. So we can use the same pattern of argument to explain why this triangle, this circle, this mug of coffee, the sun and the moon are all of them objects which are not even numbers. On the other hand, we could replace the predicate 'x is the sum of two primes' in the above argument with 'x is the sum of three primes' or 'x is the same of five primes' and so on. Then we can use the same general pattern of argument to explain why a huge variety of unrelated objects also fail to have these properties too. Such a pattern seems to be an obvious example of spurious unification.

What about Goodman's problem? Can we say that Goldbach's conjecture gives us a better explanation of the numerical data than the hypothesis that all even numbers greater than two are GRUE? (Recall that a number is GRUE if and only if it is less than or equal to $4 \cdot 10^{14}$ and the sum of two primes or greater than $4 \cdot 10^{14}$ and the sum of three primes). Perhaps we can say something like this. We have two patterns of argument:

(1) $\forall x(\mathrm{F}x \rightarrow \mathrm{G}x), \mathrm{F}a \vdash \mathrm{G}a$

(2) $\forall x(\mathrm{F}x \rightarrow \mathrm{GRUE}(x)), \mathrm{F}a,\ a \leq 4 \cdot 10^{14} \vdash \mathrm{G}a$

In more detail (2) has the form:

(2) $\forall x(\mathrm{F}x \rightarrow [(\mathrm{G}x \;\&\; x \leq 4 \cdot 10^{14}) \vee (\mathrm{H}x \;\&\; x > 4 \cdot 10^{14})]), \mathrm{F}a,\ a \leq 4 \cdot 10^{14} \vdash \mathrm{G}a$

Now in (2) the predicate H*x* – '*x* is the sum of three primes' could be replaced by any predicate at all without impairing our ability to derive the numerical data. Furthermore, by suitably replacing H*x* in (2) we can generate any conclusion at all concerning numbers greater than $4 \cdot 10^{14}$. Arguments of this form certainly seem to be less stringent than arguments of the first form. If so then arguments of type (1) will have a greater unifying power than arguments of type (2) and so that they will be more explanatory on our account.

The concept of explanation as unification also holds out some hope of explaining how there can be empirical evidence for mathematics. I argued in chapter four that the problem here is to show that mathematics can be used to *explain* physical phenomena. For if mathematics can help to explain empirical facts, then it can be confirmed by them. But how can mathematics provide us with explanations of physical facts if mathematical objects play no *causal* role in accounting for the phenomena? Given an account of explanation as unification however, we can answer that mathematics contributes to the explanation of physical phenomena by providing a framework which allows us to construct highly unified theories of the physical world. The use of the techniques of the calculus in Newton's theory of gravitation for example is essential to the unification of the phenomena achieved by that theory. Of course, the representation of physical systems by differential equations and the use of the calculus to derive the behaviour of those systems from such a

representations is now ubiquitous in science. The immense unification of our beliefs this provides should be obvious.

In the same way, the theory of complex numbers (especially complex analysis) allows us to give a unified and systematic account of such diverse phenomena as the behaviour of alternating currents in electrical circuits, aerodynamics, fluid dynamics and the properties of quantum mechanical systems. The role of complex numbers in unifying mathematics itself also has an indirect effect on the unification of our physical theories, by providing us with a powerful set of techniques for solving a wide variety of mathematical problems which arise in applications.[36]

In general, unification in mathematics contributes in a significant way to unification in the rest of science. It is not simply that mathematics allows us to *deduce* the observed phenomena from our theories. Mathematics also plays a significant role in the *explanation* of the phenomena provided by those theories, by providing us with very powerful *patterns* of argument which we can use in our physical theories to account for a wide and diverse range of phenomena. In this way mathematics contributes to the unification of our beliefs about the physical world that any explanatory theory must provide.

[36] For a more detailed discussion of the unifying role of complex numbers in mathematics and science, see [Colyvan 1991].

(1) $\forall x(Fx \rightarrow Gx), Fa \vdash Ga$

(2) $\forall x(Fx \rightarrow \mathrm{GRUE}(x)), Fa,\ a \leq 4 \cdot 10^{14} \vdash Ga$

In more detail (2) has the form:

(2) $\forall x(Fx \rightarrow [(Gx \,\&\, x \leq 4 \cdot 10^{14}) \vee (Hx \,\&\, x > 4 \cdot 10^{14})]), Fa,\ a \leq 4 \cdot 10^{14} \vdash Ga$

Now in (2) the predicate Hx – 'x is the sum of three primes' could be replaced by any predicate at all without impairing our ability to derive the numerical data. Furthermore, by suitably replacing Hx in (2) we can generate any conclusion at all concerning numbers greater than $4 \cdot 10^{14}$. Arguments of this form certainly seem to be less stringent than arguments of the first form. If so then arguments of type (1) will have a greater unifying power than arguments of type (2) and so that they will be more explanatory on our account.

The concept of explanation as unification also holds out some hope of explaining how there can be empirical evidence for mathematics. I argued in chapter four that the problem here is to show that mathematics can be used to *explain* physical phenomena. For if mathematics can help to explain empirical facts, then it can be confirmed by them. But how can mathematics provide us with explanations of physical facts if mathematical objects play no *causal* role in accounting for the phenomena? Given an account of explanation as unification however, we can answer that mathematics contributes to the explanation of physical phenomena by providing a framework which allows us to construct highly unified theories of the physical world. The use of the techniques of the calculus in Newton's theory of gravitation for example is essential to the unification of the phenomena achieved by that theory. Of course, the representation of physical systems by differential equations and the use of the calculus to derive the behaviour of those systems from such a

representations is now ubiquitous in science. The immense unification of our beliefs this provides should be obvious.

In the same way, the theory of complex numbers (especially complex analysis) allows us to give a unified and systematic account of such diverse phenomena as the behaviour of alternating currents in electrical circuits, aerodynamics, fluid dynamics and the properties of quantum mechanical systems. The role of complex númbers in unifying mathematics itself also has an indirect effect on the unification of our physical theories, by providing us with a powerful set of techniques for solving a wide variety of mathematical problems which arise in applications.[36]

In general, unification in mathematics contributes in a significant way to unification in the rest of science. It is not simply that mathematics allows us to *deduce* the observed phenomena from our theories. Mathematics also plays a significant role in the *explanation* of the phenomena provided by those theories, by providing us with very powerful *patterns* of argument which we can use in our physical theories to account for a wide and diverse range of phenomena. In this way mathematics contributes to the unification of our beliefs about the physical world that any explanatory theory must provide.

[36] For a more detailed discussion of the unifying role of complex numbers in mathematics and science, see [Colyvan 1991].

CONCLUSION: MATHEMATICS AS A SCIENCE

I would like to conclude with a few remarks on the thesis that mathematics is a science. In the introduction, I said that mathematics is a science because it is sensitive to evidence. A better way of putting it is to say that mathematics is a science because the evidence we have for our mathematical beliefs is not fundamentally different in kind to the evidence we have for our scientific beliefs.[1] The evidence we have for our beliefs about sets, numbers and functions is of exactly the same kind as the evidence we have for our beliefs about atoms, genes and galaxies. This is not to say that the evidence for our mathematical beliefs is always empirical; as we have seen, a great part of that evidence is not empirical, but purely mathematical.

One of the ways in which mathematics is often distinguished from science is by the role of *proof* in mathematics. In science, it is said, there are no deductive proofs. A scientific hypothesis can be supported by evidence, but such evidence rarely (if ever) entails the hypothesis in the way that the premises of a mathematical proof entails the conclusion. The methodology of mathematics is deductive, while the methodology of science is not.[2]

[1] Nor is the evidence we have for our scientific beliefs fundamentally any different to the evidence we have for many of everyday beliefs about more mundane matters. Science (and mathematics) is simply an extension and elaboration of ordinary rational inference. By saying that mathematical evidence is of essentially the same *kind* as scientific evidence, I mean to rule out differences that have only to with subject matter, The fact that biology treats of different kinds of things to cosmology does not mean that biological and cosmological evidence are of different kinds. In the same way, I am suggesting that the only differences between mathematics and science are differences in their subject matter. That is the sense in which mathematics is a science like any other.

[2] Here is a particularly clear expression of this view: "We use "induction" to name the form of reasoning that distinguishes the natural sciences of chemistry, meteorology, and geology from mathematical subjects such as algebra, geometry, and set theory.. Somehow or other the scientist or mathematician reaches a conclusion about the subject matter, and will seek to justify the conclusion…It is the nature of the justification...that differentiates the two cases. The mathematicians justification will be a proof. ....The point about deductive arguments, like mathematical ones, is that their assumptions or premises *entail* their conclusions. …In science, inferences from data to generalizations or to predictions typically do not entail their conclusions; they do not carry the logical inevitability of deduction." [Bird 1998, pp. 10-12].

A related distinction between science and mathematics is that in mathematics we can achieve *certainty* regarding our beliefs, while in science we cannot. For in mathematics, we can prove our claims in a way that is not possible in science. Correspondingly mathematics is *cumulative* in a way that science is not. A mathematical truth, once established is never rejected at a later date. Mathematical theories, unlike scientific theories are never overturned by rival theories that give us a better account of the phenomena, in the way that Einstein's general theory of relativity overturned Newton's theory of gravitation.[3]

We have seen however, that all three of these supposed differences between science and mathematics are illusory. Certainly *one* kind of evidence we can have for a mathematics is deductive; we may come to accept a mathematical statement on the grounds that it can be deduced from certain axioms and definitions for example. But for this to count as establishing the truth of the statement, there must be some evidence that the axioms are true. And we have seen that the evidence for the axioms is *non-deductive*. The axioms are accepted on the grounds that they can be used to explain a large body of independently acceptable statements. On the account of mathematical explanation developed in the previous chapter, axioms achieve this by providing a unification of our beliefs about the mathematical facts; they allow us to derive a large number of accepted results, by means of a few basic patterns of argument. Of course, this justification for

[3] As Bird puts it "A proof is a chain of reasoning each link of which proceeds by deductive logic. This fact lends certainty to the justification. If what we really have is a mathematical proof, then it is certain that the theorem is true so long as the premises are true. Consequently, a proof is final in that a theorem once established by proof cannot be undermined by additional data. No one is going to bring forward evidence contrary to the conclusion of Euclid's proof that there is no largest prime number. This contrasts with the scientific case....however strongly the evidence is shown to support a hypothesis, the logical possibility of the hypothesis being false cannot be ruled out. The great success of Newtonian mechanics amounted to a vast array of evidence in its favour, but this was not sufficient to rule out the possibility of its being superseded by a rival theory." [ibid. 1998, p. 11].

axioms is no different to the justification of the fundamental principles of our scientific theories. They too are accepted on the grounds that they provide us with explanations, by giving a unified and systematic account of a wide and diverse range of phenomena. In science, the evidence for the fundamental principles of our theories is the facts that they explain. Exactly the same is true of the evidence for the axioms of mathematical theories.

Of course, we will also require some independent evidence that the statements a theory explains are indeed true. In science, this is usually achieved by observation and experiment. Empirical evidence may also play this role in mathematics. Some kinds of deductive justification may also serve this function – calculation for example, or more generally, confirmation-techniques of various kinds. The independent evidence for the statements which confirm the axioms of a mathematical theory may also be of various non-deductive kinds; induction, analogy, verification of a consequence and so on.

On my account, these all conform to the same basic pattern; the evidence which supports a statement is explained by it. Instances of a universal statement which are explained by it provide us with *inductive* evidence. Statements concerning observable features of the world which are explained by a mathematical statement provide us with *empirical* evidence for it. More generally, any consequence of a mathematical statement which is also explained by it provides us with evidence for that statement. And in each case, the better the explanation, the stronger the evidence.

Since the evidence for the premises of a mathematical proof is at root non-deductive, proof in mathematics cannot provide us with absolute certainty. A proof in a developing theory may implicitly appeal to premises which are dubious or even false. Proofs from axioms, though more certain (since the premises are better established) cannot

provide us with conclusive evidence, since the evidence for the axioms, being non-deductive, does not establish them conclusively.

The illusion that there is a difference between mathematics and science in this respect is generated by the fact that mathematics tends to throw away its props. Mathematicians have a tendency to leave out all the non-deductive evidence which led to a theorem when they publish a proof of it. The presentation of mathematics in textbooks tends to leave out the long and painful historical process which led to the theory and simply presents the final polished deductive structure. These practices tend to create the illusion that in mathematics we simply begin from self-evident principles and then deduce theorems from them. But we have seen that the real history of mathematics presents a very different picture of evidence in mathematics.[4]

Consider also the way in which a derivation can turn into a proof. A body of mathematical results are initially supported by a heterogeneous collection of empirical and non-deductive evidence. A systematic theory (axiomatization) is then proposed to account for all these results. The evidence for the axioms of the theory is that they can be used to explain those results. At this point the axioms are supported by the lower level results and the latter are supported by induction and observation. But once the axioms have become well supported by this kind of evidence, we throw away the scaffolding and the derivations which once supported the axioms turn into proofs. We are left with an apparently free standing structure, one in which the theorems are justified by the axioms and the axioms appear to rest on nothing at all, except perhaps themselves.

[4] Reuben Hersh has described some of the ways in which mathematics presents a false image of itself in 'Mathematics has a Front and Back'. [Hersh 1991].

We have also seen how accepted mathematical statements may later come to be rejected on the basis of counter-examples, or in the light of an improved theory. Recall, for example, the case of Leibniz' and Euler's claims about the sums of divergent series or Cauchy's assertion that a convergent series of continuous functions is continuous. Mathematical theories may also be superseded by others, perhaps more rigorous or systematic. It is certainly rare for a mathematical theory to be simply rejected as false. What more commonly happens that the theory is reinterpreted, as in the case of Euclidean geometry, or subsumed under a more general theory, as occurred with Hamilton's theory of quaternions.[5]

But this is often what happens in science too. Kepler's laws of planetary motion were superseded by Newton's theory of gravitation, by showing that those laws could be derived from the more general theory as a special case. In the same way, although Newtonian mechanics has been superseded by the theory of relativity, it has not been rejected in its entirety as simply false. Rather, it has been subsumed under the theory of relativity as a special case; an approximation to the truth which holds when velocities are small in comparison to the speed of light.

Is mathematics nonetheless *more* certain than empirical science? Perhaps it is so. The reason, I suspect, has to do with the nature of the objects of mathematics. In empirical science, we have an independent, perceptual access to the objects of theories; we can see them, touch them and measure them. We have no such perceptual access to the objects of mathematics; a number is not something you can see or touch. This does not mean that we can have no empirical evidence for the existence of such objects, just as the fact that many of the objects studied by science are not directly observable does not mean we cannot have

[5] See [Crowe 1985, 1988, p. 263, Fisher 1966]

empirical evidence for them. For we can have evidence for statements which refer to objects of a certain kind, even though we have no direct access to those objects.[6] (Think of our beliefs about dinosaurs for example). The independent access we have to physical objects makes our theories about them highly liable to falsification by recalcitrant experience. Mathematical objects on the other hand are known mostly through our theories of them.[7] We have no independent access to those objects, and so our mathematical theories are far less liable to falsification than empirical theories. That, I would suggest is why mathematics seems to provide us with greater certainty. Of course this does not imply that our mathematical knowledge is infallible. Although we cannot refute a mathematical theory by examining its objects under a microscope, we have seen that the evidence we have for those theories is certainly not of a kind which makes them immune from revision.

Despite appearances then, there is no strong disanalogy between science and mathematics which respect to proof. Certainly there are no proofs in science in the sense of arguments which provide us with absolute certainty. But there is no such thing as proof in this sense in mathematics either. Another way of stating this same fact is to say that there is proof in a science, in exactly the same sense that there is proof in mathematics.

A large part of the activity of the working scientist consists in showing that the currently accepted theory can be used to predict and explain new phenomena. But this is precisely what a mathematician is doing when he or she proves a new theorem; using a well established theory to derive a new result.

[6] This, perhaps, is the lesson of Frege's context principle; we do not 'grasp' numbers, but this does not mean we can know nothing about them – our knowledge of numbers consists in our knowledge of their laws, that is, in our knowledge of propositions referring to them.
[7] See also [Hart 1979, pp. 61-2].

So in the sense that there is proof in mathematics, there is equally proof in science. Newton *proved* Kepler's laws of motion in exactly the same sense that Dedekind proved the associative law of addition. Scientific proofs are usually called *explanations*. The fact that we speak of *proof* in mathematics does not alter the fact that a proof of a theorem is really an explanation of it. A proof in mathematics is a derivation of a statement from an accepted mathematical theory, just as a scientific explanation may be a derivation of a statement from a physical theory.

Mathematicians prefer proof to inductive confirmation of a conjecture, because a proof gives us a better explanation of why the conjecture is true. In exactly the same way, observation and experiment may confirm a physical hypothesis, but we get a better explanation of it by showing how we can derive it from a system of scientific laws or theory. Kepler's laws were well-confirmed by observation, but Newton's derivation of those laws from his theory provides us with an explanation of them. This kind of explanation is analogous to proof in mathematics. In both cases, explanation is achieved by unification; we look for a few basic principles which allow us to give a systematic account of as many of the facts as possible.

Explanation by unification is certainly very common in science. We can explain the physical world by showing how a wide and diverse range of phenomena can be deduced from a few fundamental laws. This may not be the only kind of explanation in the empirical sciences[8], but I suggest it is the rule in mathematics.

[8] A different kind of scientific explanation consists in describing a mechanism; a causal process underlying the phenomena to be explained. [see Salmon 19xx, p. y]. This kind of explanation is unavailable in mathematics. Again, I would suggest that this difference between science and mathematics can be accounted for entirely in terms of differences in their subject matters.

The historical development of mathematics reveals a continual drive towards ever increasing unification of the field. At any stage of mathematical inquiry, there is collection of accepted results or statements, and a set of problems or questions. The accepted results will be supported by evidence of various kinds, which we can represent as a set of arguments or justifications for those results. In the very early stages of mathematics, the accepted statements may be empirical, a set of observed regularities in the relations between shapes of physical objects and the ways in which they can be grouped or counted perhaps. The problems and questions are likewise empirical; concerned with solving various practical problems of measurement and distribution. We then introduce some simple rules and principles of arithmetic and geometry which are justified empirically or pragmatically, on the grounds that they can be used to explain those regularities or solve those practical problems. This gives us a basis of mathematical facts on which the process of unification can set to work.

Some of the statements we come to accept in this way may later be given a deductive justification, by showing how they can deduced from other accepted statements, perhaps by means of certain general principles, which may or may not be explicitly stated. Recognising those principles and adding them to our store of accepted statements results in an improvement in the unification of our mathematical beliefs. At the same time, adding those statements may raise new questions.

These deductive justifications may provide us only with a set of confirmation-techniques (rules for calculation for example) for answering questions. New concepts and statements may then be introduced into mathematics on the grounds that they provide problem-solutions to those questions. On my account, the evidence for these new statements is that they explain many of the prior results by providing a general pattern of argument which we can use to derive a large number of related results. Again, what this achieves is an increase in the unification of mathematics.

As mathematics continues to develop, we see the creation of increasingly unified theories of the mathematical facts discovered so far. Now, there is an ambiguity in the term 'theory' as applied in mathematics. There is a difference between for example, theories in the sense of number theory, set theory, analysis, Euclidean geometry and so on and theories in the sense of the theory of the distribution of primes, the theory of partial differentiation, the theory of equations, solid geometry and so on. The former are deductively closed, systematic, axiomatized sets of sentences about a certain fundamental kind of mathematical structure. These correspond to fundamental theories in science such as quantum mechanics and general relativity. The latter are collections of results and theorems about a more specific area of mathematical research. These correspond to physical theories such as electrical circuit theory, fluid dynamics, optics and so on. This kind of theory is often embedded in a more general theory of the first kind, but need not be.

Theories of this second kind can contribute to the unification of mathematics in various ways, by introducing new general principles or concepts which allow us to give a more systematic treatment of the field in question; by replacing confirmation-techniques with problem-solutions; by providing a more rigorous or general treatment of the field by means of new basic principles and definitions; and so on. In all cases, the justification for these additions to mathematics is that they provide explanations of the body of prior results by increasing the unification of the field in various different ways.[9] Of course, they also

[9] In this way I hope to have shown how the patterns of change in mathematical practices described by Kitcher can all be seen as being justified in essentially the same way.

serve a second goal of mathematics; that of discovering new mathematical truths by allowing us to derive new statements, not previously accepted.

Theories in mathematics always begin as theories of the second kind. We have seen that before the nineteenth century, there were no theories in the first kind, except perhaps for Euclidean geometry. Arithmetic, algebra, calculus and set theory all began as theories in the second sense; none of them were embedded in a rigorous, axiomatic theory. Similar remarks apply to many theories in the rest of science of course. Nonetheless, just as in the case of science, theories of this kind tend to gradually become incorporated into more fundamental theories of the first kind. Just as in physics, the theory of heat became incorporated into atomic theory and mechanics and the latter becomes incorporated into quantum mechanics; so in mathematics, elementary arithmetic gets incorporated into number theory and the latter comes to be incorporated into set theory. In the same the various different kinds of group structures which had been studied by mathematicians (permutation and substitution groups, transformation groups and so on) all became incorporated into abstract group theory, which in its turn came to be subsumed under set theory.[10] The general pattern is the same in both science and mathematics; disparate results from different theories get incorporated into increasingly unified and fundamental theories.

This drive towards unified theories in mathematics is not different to the search for unified theories of the fundamental forces of nature in physics. In mathematics, the drive towards unification seems now to have reached its peak in axiomatic set-theory, from which all known braches of mathematics can be derived. But current set theory is not the

[10] See [Kline 1972, pp. 772-791, 1136-1156].

end of this story. We know that there are many questions it leaves unanswered. Who knows where the search for explanatory unification in mathematics will lead us next?

I argued in chapter one, that in mathematics, there are no foundations, only better and worse kinds of evidence. We can now, I hope, see more clearly what this comes to. The distinction between better and worse kinds of evidence in mathematics corresponds to the distinction between better and worse kinds of explanation. We have better evidence for a mathematical statement when we have a better explanation of it, and on my account, the better explanation is the one that contributes more to the unification of our mathematical beliefs.

I have argued that mathematics is a science on the grounds that this is the view which makes the most sense of the practice of mathematics. If we look at what mathematicians actually do we find that they present evidence of various kinds for mathematical statements, attempt to construct explanatory, unified theories of the known mathematical facts, attempt to show that the resources provided by those theories can be used to derive new results, struggle with problems created by tensions within and between different theories, propose hypotheses and conjectures, find counter-examples to them, modify their beliefs in the light of new evidence and theories and so on. All of this is no different to the activity of scientists working in fields such as physics, chemistry and biology and other sciences.

Many of the philosophers whose views on the epistemology of mathematics I have examined here can be seen as advancing the thesis that mathematics is a science, albeit in different ways, corresponding to their differing conceptions of the nature of science in

general. Frege, for example, often referred to mathematics as a science[11], but he operated with a Cartesian or rationalist conception of science. On Frege's account, the theorems of a branch of mathematics are justified by deducing them form first principles. But in this respect, mathematics is no different to any other science. What distinguishes mathematics from the other science is not that justification in mathematics is deductive, while justification in the empirical sciences is not, for in Frege's view *all* justification is deductive; every science has its first principles from all the truths of that branch of science can be deduced. What distinguishes the science of mathematics from sciences such as physics and chemistry is the mode of justification of the first principles involved; the principles of the empirical sciences are justified by observation and perception, while the principles of mathematics are justified either by pure logic (in the case of arithmetic and analysis) or by pure intuition (in the case of geometry).

Hilbert's formalist account of mathematics, by contrast, can be seen as an expression of a very different view of science; the operationalist or instrumentalist view favoured by the positivists. On this view, the only meaningful propositions – certainly the only ones whose truth it is the business of science to establish – are those that can be directly verified or falsified by observation. Scientific theories that refer to unobservable objects and mechanisms are to be thought of as no more than useful instruments for deriving statements about observables. Analogously, on Hilbert's account, the only meaningful propositions are those that can be verified or falsified by finitary computation. Mathematical theories that go beyond what is verifiable in this way, by treating of

[11] For example: 'It is applicability alone that raises arithmetic from the rank of a game to that of science'. [*Grundgesetze*, vol. II, §91]. See also [*Grundlagen*, p. II] were Frege describes our inability to say what the number 1 is as a scandal to the science of mathematics.

completed infinite totalities, are no more than useful instruments for deriving true finitary propositions.[12]

In the same way, Lakatos's account of mathematics as a science on the other hand, is informed by Popper's falsificationist model of science, while Kitcher gives an improved account by making use of a more realistic picture of the development of science, based on the concept of the evolution of scientific or mathematical practices.

We saw in chapter four how Quine also arrives at the conclusion that mathematics is a science (or more accurately an integral part of science) by reflecting on the inadequacies of reductionist accounts of scientific theories, according to which the theoretical portions of out theories can be eliminated in favour of statements referring only to what is directly observable. We also saw how the problems with Quine's account of the epistemology of mathematics are connected with problems with his overall account of science in general; in particular problems with his hypothetico-deductive account of scientific evidence. Similar remarks apply to Lakatos's account, which inherits many of the problems with falsificationism in the philosophy of science more generally.

I think that the general pattern here is clear. The better account of mathematics corresponds to the better account of science in general. They key to progress in the epistemology of mathematics is to improve our understanding of the nature of science in general. Equally, our understanding of science will be improved by a closer examination of mathematics.

No doubt there are a great many problems with the account of mathematical evidence which I have offered. I hope to have at least pointed out the right way to approach the solution of such problems; by confronting the account with mathematical practice and

[12] See [Giaquinto 1983, pp. 125-7].

methodology, as revealed in its historical development. I hope also to have shown that many problems and issues in the epistemology of mathematics are fundamentally the same as problems and issues in the philosophy of science more generally – especially those connected with the related concepts of evidence and explanation.

REFERENCES

ARNAULD, A. AND NICOLE, P. 1872: *The Port Royal Logic*. Translation by Thomas Spencer Baynes, 7th edition, Edinburgh: Oliver & Boyd, 1872. Also translated as *The Art of Thinking: Port Royal Logic*, by J. Dickoff and P. Jones, Indianapolis, Bobbs-Merrill, 1964.

ASPRAY, W. AND KITCHER P. (eds.) 1988: *History and Philosophy of Modern Mathematics*. Minnesota Studies in the Philosophy of Science. Vol. 11. University of Minnesota Press, 1989.

AYER, A. J. 1936: 'The *A Priori*'. In Ayer 1946, pp. 71-87. Excerpted in Benacerraf and Putnam 1983, pp. 315-328.

__________ 1946: *Language, Truth and Logic*. Second Edition, Penguin Books, 1946.

BELNAP, N. AND STEEL, T. 1976: *The Logic of Questions and Answers*. New Haven: Yale University Press, 1976.

BENACERRAF, P. 1965: 'What Numbers Could Not Be'. *Philosophical Review*, vol. 74, 1965. pp. 47-73. Reprinted in Benacerraf and Putnam 1983, pp. 272-294.

__________ 1973: 'Mathematical Truth'. *Journal of Philosophy*, vol. 70, no. 19, 1973, pp. 661-679. Reprinted in Benacerraf and Putnam 1983, pp. 403-420 and in Hart 1996, pp. 14-30.

BENACERRAF, P. AND PUTNAM, H. (eds.) 1983: *Philosophy of Mathematics. Selected Readings*. 2nd Edition, Cambridge: Cambridge University Press, 1983.

BIRD, A. 1998: *Philosophy of Science*. Fundamentals of Philosophy series, London: UCL Press, 1998.

BOLZANO, B. 1817: *Rein analytischer Beweis*. Prague: Gottloeb Hass, 1817. Reprinted in *Ostwald's Klassiker der exacten Wissenschaften*, vol. 153, 1905, pp. 3-43.

BONOLA, R. 1955: *Non-Euclidean Geometry – A critical and historical study of its development*. New York: Dover Press, 1955.

BOOLOS, G. 1971: 'The iterative conception of set'. *Journal of Philosophy*, vol. 68, 1971, pp. 215-32. Reprinted in Benacerraf and Putnam 1983, pp. 486-502.

__________ 1995a: 'The Consistency of Frege's Foundations of Arithmetic'. In Demopoulous 1995.

BOLYAI, J. 1832: 'On Non-Euclidean Geometry'. Translated by H. P. Manning in Smith 1959, pp. 375-388.

__________ 1995b: 'The Standard of Equality of Numbers'. In Demopoulous 1995.

BORGES, J. L. 1964: 'The Fearful Sphere of Pascal'. *Labyrinths*, Penguin 1964.

BROUWER, L. E. J. 1913: 'Intuitionism and Formalism'. *Bulletin of the American Mathematical Society*, vol. 2. Trans. A. Dresden. Reprinted in Benacerraf and Putnam 1983, pp. 77-89.

__________ 1949: 'Consciousness, Philosophy, and Mathematics'. In *Proceedings of the Tenth International Congress of Philosophy*, vol. 1, pt. 2. Reprinted in Benacerraf and Putnam 1983, pp.90-96.

BROWN, J. R. 1990: 'Proof and Truth in Lakatos's Masterpiece'. *International Studies in the Philosophy of Science*, vol. 4, no.2, 1990, pp. 117-130.

__________ 1997: 'Proofs and Pictures'. *British Journal for the Philosophy of Science*, vol. 48, 1997, pp. 161-180.

__________ 1999: *Philosophy of Mathematics. An introduction to the world of proofs and pictures*. London and New York: Routledge, 1999.

BURALI-FORTI, C. 1897: 'A Question on Transfinite Numbers'. Reprinted in van Heijenoort 1967, pp. 105-12.

BURTON, D. M. 1998: *The History of Mathematics. An Introduction*. Allyn and Bacon, 4th Edition, 1998.

CAIRNS-SMITH, A. G. 1985: *Seven Clues to the Origin of Life – A Scientific Detective Story*. Cambridge: Cambridge University Press, 1985.

CALDWELL, C. 2000: 'An Amazing Prime Heuristic'. Preprint available from http://www.utm.edu/~caldwell/preprints/Heuristics.pdf

CANTOR, G. 1878: 'Ein Beitrag zur Mannigfaltigkeislehre'. *Journal für due reine und angewandte Mathematik*, vol. 84, 1878, pp. 242-58.

__________ 1883: 'Über unendliche lineare Punktmannichfaltigkeiten, V'. *Mathematische Annalen*, vol. 21, 1883, pp. 545-86.

__________ 1895-7: 'Beiträge zur Begründung der transfiniten Mengenlehre'. *Mathematische Annalen*, vol. 46, 1895, pp. 481-512, Vol. 49, 1897, pp. 207-246. Translated in Cantor 1955.

__________ 1932: *Gesammelte Abhandlungen*. E. Zermelo (ed.). Berlin: Springer, 1932.

__________ 1955: *Contributions to the Founding of the Theory of Transfinite Numbers*. Translated with an introduction by P. E. B Jourdain. New York: Dover Press, 1955.

CARNAP, R. 1931: 'Die logizistiche Grudlegung der Mathematik". *Erkenntnis*, vol. 1. Translated as 'The Logicist Foundations of Mathematics' in Benacerraf and Putnam 1983, pp. 41-52.

CAUCHY, A. L. 1811: 'Recherches sur les polyèdres', *Journal de l'Ecole Polytechnique*, vol. 9, 1813, pp. 68-86.

CHEYNE, C AND PIGDEN. C. 1996: 'Pythagorean Powers or A Challenge to Platonism'. *Australasian Journal of Philosophy*, vol. 74, no. 4, 1996, pp. 639-45.

CHIHARA, C. 1990: *Constructibility and Mathematical Existence*. Oxford: Oxford University Press, 1990.

CHURCH, A. 1949: 'Review of Ayer'. *Journal of Symbolic Logic*, vol. 14, 1949, pp. 52-53.

__________ 1956: *Introduction to Mathematical Logic, Part I*. Princeton: Princeton University Press, 1956.

CHURCHHOUSE, R.F AND GOOD, I.J: 1968: 'The Riemann Hypothesis and Pseudorandom Features of the Möbius Sequence.' *Mathematics of Computation*, vol. 22, 1968, pp. 857-64.

COHEN, P. 1963-4: 'The Independence of the Continuum Hypothesis, I and II'. *Proceedings of the National Academy of Sciences*, vol. 50-1, 1963-4. pp. 1143-8,105-10.

COLYVAN, M. 1998a: 'Is Platonism a Bad Bet?'. Discussion of Cheyne and Pigden 1996. *Australasian Journal of Philosophy*, vol. 76, no. 1, pp. 115-19.

__________ 1998b: 'In Defence of Indispensability'. *Philosophia Mathematica*, vol. 6, 1998, pp. 39-62.

__________ 1999: 'Confirmation Theory and Indispensability'. *Philosophical Studies*, vol. 96, no. 1, 1999, pp. 1-19.

CRAIG, W. 1953: 'On axiomatizability within a system'. *Journal of Symbolic Logic*, vol. 18, no. 1, 1953, pp. 30-2.

__________ 1956: 'Replacement of auxiliary expressions'. *Philosophical Review*, vol. 65, no. 1, 1956, pp. 38-55.

CROSSLEY, J. N. 1980: *The Emergence of Number*. Steel's Creek: Upside Down A Book Company, 1980.

CROWE, M. J. 1985: *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System*. New York: Dover Press, 1985.

__________ 1988: 'Ten Misconceptions about Mathematics and Its History'. In Aspray and Kitcher (eds.) 1988, pp. 260-77.

DAVIS, P.J AND HERSH, R. 1981: *The Mathematical Experience*. Harvester Press, 1981.

DAWKINS, R. 1986: *The Blind Watchmaker*. Penguin Books, 1986.

DEDEKIND, R. 1872: *Stetigkeit und irrationale Zahlen*. Translated as 'Continuity and Irrational Numbers' by W. Beman in Dedekind 1963.

__________ 1888: *Was sind und was sollen die Zahlen?* Translated as 'The Nature and Meaning of Numbers' by W. Beman in Dedekind 1963.

__________ 1963: *Dedekind's Essays on the Theory of Numbers*. Edited by W. Beman. New York: Dover Press, 1963.

DEMOPOULOS, W. (ed.) 1995: *Frege's Philosophy of Mathematics*. Cambridge, Mass.: Harvard University Press, 1995.

DESCARTES, R. 1628: 'Rules For The Direction of The Mind'. In *Descartes Philosophical Writings*. Translated by E. Anscombe and P.T Geach. London: Thomas Nelson & Sons, 1954.

__________ 1897-1913: *Oeuvres*. Adam and Tannery (eds.) Paris: 1897-1913.

DETLEFSEN, M. 1988: 'Fregean Hierarchies and Mathematical Explanation'. *International Studies in the Philosophy of Science*, vol. 3, no. 1, 1988, pp. 97-116.

__________ 1992: *Proof and Knowledge in Mathematics*. (ed.) London: Routledge, 1992.

__________ 1992a: 'Brouwerian Intuitionism'. In Detlefsen (ed.) 1992.

__________ 1992b: 'Poincaré against the logicians'. *Synthese*, vol. 90, 1992, pp. 349-78.

DEVLIN, K. 1988: *Mathematics: The New Golden Age*. Penguin Books, 1988.

DUHEM, P. 1906: *The Aim and Structure of Physical Theory*. Princeton University Press, 1954.

DUMMETT, M. 1975: 'The Philosophical Basis of Intuitionistic Logic'. In *Logic Colloquium 1973. Proceedings of the Logic Colloquium. Studies of the Logic and the Foundations of Mathematics*, vol. 80. Reprinted in Dummett 1978, Benacerraf and Putnam 1983 and Hart 1996.

__________ 1978: *Truth and Other Enigmas*. Harvard University Press, 1978.

__________ 1991: *Frege Philosophy of Mathematics*. Harvard University Press, 1991.

EDWARDS, H.M. 1974: *Riemann's Zeta Function*. New York, London: Academic Press, 1974.

EINSTEIN, A. 1917: *Relativity: The Special and the General Theory*. Translated by R.W. Lawson. New York: Crown, 1961.

EULER, L. 1750: 'Elementa Doctrinae Solidorum'. *Novi Commentarii academiae scientiarum Petropolitanae*, 1758, vol. 4, pp. 109-140.

________ 1770: *Elements of Algebra*. Translated by J. Hewlett, Springer-Verlag, 1984.

________ 1911-36: *Opera Omnia*. Leipzig: Teubner, 1911-36.

FEIGL, H. AND SELLARS, W. (eds.) 1949: *Readings in Philosophical Analysis*. Appleton-Century-Crofts, 1949.

FIELD, H. 1980: *Science Without Numbers*. Oxford: Basil Blackwell, 1980.

__________ 1985: 'On Conservativeness and Incompleteness'. *Journal of Philosophy*, vol. 81, pp. 239-60. Reprinted in Field 1989, pp. 125-146.

__________ 1989: *Realism, Mathematics and Modality*. Blackwell, Oxford 1989.

FISHER, C. 1966: 'The Death of a Mathematical Theory: A Study in the Sociology of Knowledge'. *Archive for the History of the Exact Sciences*, vol. 3, 1966, pp. 137-59.

FORDER, H. G. 1927: *The Foundations of Euclidean Geometry*. New York: Dover Press, 1927.

FRANKLIN, J. 1987: 'Non-deductive Logic in Mathematics'. *British Journal for the Philosophy of Science*, vol. 38, 1987, pp. 1-18.

FREGE, G. 1879: *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Nebert, 1879. Translated as "Begriffsschrift, a Formal Language, Modelled upon that of Arithmetic, for Pure Thought" by S. Bauer-Mengelberg in van Heijenoort 1967.

__________ 1884: *Die Grundlagen Der Arithmetik*. Translated as *The Foundations of Arithmetic* by J. L. Austin. Oxford: Basil Blackwell, 1978.

__________ 1893-1903. *Grundgesetze der Arithmetik*, 2 vols. Translated in *The Basic Laws of Arithmetic: Exposition of the System* by M. Furth, University of California Press, 1964.

__________ 1979: *Gottlob Frege: Posthumous Writings*. Translated by P. Long, R. White and R. Hargreaves. London: Blackwell, 1979.

FRIEDMAN, M. 1974: 'Explanation and Scientific Understanding'. *Journal of Philosophy*, vol. 71, 1974, pp. 5-19.

GALILEO, G. 1952. *Dialogues Concerning Two New Sciences*. New York: Dover Press, 1952.

GAMOW, G. 1966: *Thirty Years That Shook Physics. The Story of Quantum Theory*. London: Heinemann, 1966.

GAUSS, C. F. 1827: 'Disquisitiones Generales circa Superficies Curvas'. *Werke*, 4, 217-58. Translated as *General Investigations of Curved Surfaces*. Raven Press, 1965. Reprinted in Smith 1959, pp. 463-75.

GENTZEN, G. 1943: 'Provability and Non-provability of Restricted Transfinite Induction in Elementary Number Theory'. Reprinted in Gentzen 1969.

__________ 1969: *The Collected Papers of Gerhard Gentzen*. Ed. M. Szabo, Amsterdam: North-Holland, 1969.

GETTIER, E. L. 1963: 'Is Justified True Belief Knowledge?'. *Analysis*, 23, 1963, pp. 121-3.

GIAQUINTO, M. 1983: 'Hilbert's Philosophy of Mathematics'. *British Journal for the Philosophy of Science*, 34, 1983, 114-132.

__________ 1992: 'Visualizing as a Means of Geometrical Discovery'. *Mind and Language*, vol. 7, No. 4, 1992.

__________ 1993: 'Visualizing in Arithmetic'. *Philosophy and Phenomenological Research*, vol. 53, no. 2, 1993, pp. 385-96.

__________ 1994: 'Epistemology of Visual Thinking in Elementary Real Analysis'. *British Journal for the Philosophy of Science*, 45, 1994, pp. 789-813.

GLYMOUR, C. 1980a: *Theory and Evidence*. Princeton: Princeton University Press, 1980.

__________ 1980b: 'Discussion: Hypothetico-Deductivism is Hopeless'. *Philosophy of Science*, vol. 47, 1980, pp. 322-5.

GÖDEL, K. 1931: 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I.'. Translated as 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems' by J. van Heijenoort in van Heijenoort 1967.

__________ 1938: 'The consistency of the axiom of choice and of the generalized continuum hypothesis'. In Gödel 1990, pp. 26-7.

__________ 1939: 'Consistency proof for the generalized continuum hypothesis'. In Gödel 1990, pp. 28-32.

__________ 1944: 'Russell's Mathematical Logic'. In *The Philosophy of Bertrand Russell*. P. A. Schilpp (ed.) Northwestern University Press, 1944. Reprinted in Benacerraf and Putnam 1983, pp. 447-69.

__________ 1947: 'What is Cantor's Continuum Problem?'. In Benacerraf and Putnam, 1983, pp. 470-485.

__________ 1990: *Collected Works. Volume II*. ed. S. Feferman *et al*. New York: Oxford University Press.

GOLDMAN, A. 1967: 'A Causal Theory of Knowing'. *Journal of Philosophy*, vol. 64, no. 12, 1967, pp. 357-372.

GOODMAN, N. 1983. *Fact, Fiction and Forecast*. 4th Edition, Harvard University Press, 1983.

GRÜNBAUM, A. 1964: *Philosophical Problems of Space and Time*. London: Routledge and Kegan Paul, 1964.

HAACK, S. 1978: *Philosophy of Logics*. Cambridge: Cambridge University Press. 1978.

HALE, B. 1987. *Abstract Objects*. London: Basil Blackwell, 1987.

__________ 1994. 'Is Platonism Epistemologically Bankrupt?'. *Philosophical Review*, vol. 103, no. 2, 1994.

HALE, S. 1988: 'Space-time and the abstract/concrete distinction'. *Philosophical Studies*, vol. 53, pp. 85-120.

HARDY, G. H. 1929: 'Mathematical Proof'. *Mind*, vol. 37, no. 149, 1929, pp.1-25.

HARDY, G. H AND WRIGHT, E. M 1979: *An Introduction to the Theory of Numbers*. Oxford: Oxford University Press, 1979.

HARMAN, G. 1973: *Thought*. Princeton University Press, 1973.

HART, W. D. 1977: Review of Mark Steiner's *Mathematical Knowledge*. (Steiner 1975). *Journal of Philosophy*, 74, 1977, pp. 118-29.

__________ 1979: 'Access and Inference'. *Proceedings of the Aristotelian Society*, suppl. vol. 53, 1979, pp. 153-65. Reprinted in Hart 1996, pp. 52-62.

__________ 1996: (ed.) *The Philosophy of Mathematics*. Oxford: Oxford University Press, 1996.

HAZEN, A. 1985: Review of C. Wright's *Frege's Conception of Numbers as Objects*. (Wright 1983). *Australasian Journal of Philosophy*, vol. 63, 1985, pp. 251-54.

HEATH, T. L. 1956: *Euclid's Elements*. 2nd Edition. New York: Dover Press, 1956.

HELLMAN, G. 1989: *Mathematics Without Numbers*. Clarendon Press, 1989.

HEMPEL, C. 1945a: 'Studies in the Logic of Confirmation'. *Mind*, 54, 1945, pp. 213-214.

__________ 1945b: 'On the Nature of Mathematical Truth'. *American Mathematical Monthly*, vol. 52. Reprinted in Benacerraf and Putnam 1983, pp. 377-394.

__________ 1965: *Aspects of Scientific Explanation*. London: Macmillan, 1965.

__________ 1965a: 'Empiricist Criteria of Cognitive Significance: Problems and Changes'. In Hempel 1965.

__________ 1966: *Philosophy of Natural Science*. Prentice Hall, 1966.

HERSH, R. 1979: 'Some Proposals for Reviving the Philosophy of Mathematics'. *Adventures in Mathematics*, vol. 31, 1979, p. 43.

__________ 1991: 'Mathematics has a front and back'. *Synthese*, vol 88, 1991, pp. 127-33.

HEYTING, A. 1931: 'Die Intuitionistische Grudlegung der Mathematik'. *Erkenntnis*, vol. 2. Translated as 'The Intuitionistic Foundations of Mathematics' by E, Putnam and G.J Massey in Benacerraf and Putnam 1983, pp.52-61.

HILBERT, D. 1899: *Grundlagen der Geometrie*. Leipzig: Teubner, 1899. Translated as Hilbert 1971.

__________ 1926: 'Über das Unendliche'. *Mathematische Annalen*, vol. 95. Translated as 'On the Infinite' by E. Putnam and G. J Massey in Benacerraf and Putnam 1983, pp. 183-201.

__________ 1971: *The Foundations of Geometry*. Translation of Hilbert 1899 (10th edition) by L. Under. La Salle: Open Court, 1971.

JACOB, H. R 1987: *Geometry*. W. H Freeman and Co., 1987.

JAMES, R.D. 1949: 'Recent Progress in the Goldbach Problem'. *Bulletin of the American Mathematical Society*, vol. 55, pp. 246-60.

JOURDAIN, P.E.B 1955: Introduction to *Contributions to the Founding of Transfinite Set Theory* [Cantor 1955].

JOWETT, B. 1949: Translation of Plato's *Meno*. New York: Liberal Arts Press, 1949.

KAC, M. 1959: *Statistical Independence in Probability, Analysis and Number Theory*. Mathematical Association of American. Carus Mathematical Monographs, no. 12. New York: Wiley, 1959.

KANIGEL, R. 1991: *The Man Who Knew Infinity: A life of the genius Ramanujan*. New York: Washington Square Press, 1991.

KANT, I. 1781: *Kritik der reinen Vernunft*. Riga; 2nd ed. 1787. Translated as Kant 1881.

__________ 1881: *Critique of Pure Reason*. Translation of Kant 1781, (2nd ed.) by F.M Müller. London: Macmillan, 1881.

KOLATA, G.B. 1974: 'Riemann Hypothesis: elusive zeros of the zeta function'. *Science*, vol. 185, 1974, pp. 429-31.

KITCHER, P. 1973: 'Fluxions, Limits and Infinite Littlenesse'. *Isis*, vol. 64, 1973, pp. 33-49.

__________ 1975: 'Bolzano's Ideal of Algebraic Analysis'. *Studies in the History and Philosophy of Science*, vol. 6, 1975, pp. 229-267.

__________ 1976: 'Explanation, Conjunction and Unification'. *Journal of Philosophy*, vol. 73, pp. 207-12.

__________ 1977: 'On the Uses of Rigorous Proof' – Review of *Proofs and Refutations*. *Science*, vol. 196, May 1977, pp. 782-3.

__________ 1979: 'Frege's Epistemology'. *Philosophical Review*, vol. 88, no.2, 1979, pp. 235-62.

__________ 1980: 'Arithemetic for the Millian'. *Philosophical Studies*, vol. 37, 1980, pp. 215-36.

__________ 1981: 'Explanatory Unification'. *Philosophy of Science*, vol. 48, 1981, pp. 507-531.

__________ 1984: *The Nature of Mathematical Knowledge*. Oxford University Press, 1984.

KITCHER P. AND SALMON. W. 1987: 'Van Fraassen on Explanation'. *Journal of Philosophy*, vol. 84, 1987, pp. 315-30.

KLINE, M. (ed.) 1969: *Mathematics in the Modern World.* San Francisco: Freeman, 1969.

__________ 1972: *Mathematical Thought from Ancient to Modern Times*. Oxford University Press. 1972.

KRIPKE, S. 1980: *Naming and Necessity*. Harvard University Press, 1980.

KUBILIUS, J. 1964: *Probabilistic Methods in the Theory of Numbers*. American Mathematical Society. Mathematical Monographs, no. 11, 1964.

KUHN, T. S. 1970: *The Structure of Scientific Revolutions*. Chicago University Press, 1970.

LAKATOS, I. 1963: 'Proofs and Refutations'. *British Journal for the Philosophy of Science*, vol. 14, no. 53, pp. 1-342. Also published, with additional appendices, as Lakatos 1976.

__________ 1970: 'Falsification and the Methodology of Scientific Research Programmes' in Lakatos and Musgrave (eds.) *Criticism and the Growth of Knowledge*. Cambridge University Press, 1970.

__________ 1976: *Proofs and Refutations*. Cambridge University Press, 1976.

__________ 1978: *The Methodology of Scientific Research Programmes*. (ed. J Worrall & G. Currie). Cambridge University Press, 1978.

LAKATOS, I. AND MUSGROVE, A. (eds.) 1970: *Criticism and the Growth of Knowledge*. Cambridge University Press, 1970.

LE CORBEILLER, P. 1954: 'The Curvature of Space'. *Scientific American*. Nov. 1954. Reprinted in Kline (ed.) 1969.

LEIBNIZ, G. 1849-63: *Mathematische Schriften*. Edited by C. Gerhardt. Halle, 1849-63.

LEWIS, D. 1996: 'Elusive Knowledge'. *Australasian Journal of Philosophy*. vol. 74, no. 4, 1996.

LINSKY, L. (ed.) 1952: *Semantics and the Philosophy of Language*. University of Illinois Press, 1952.

LIPTON, P. 1993: *Inference to the Best Explanation*. London: Routledge, 1993.

LOBACHEVSKY, N.I. 1855: *Pangeometry*. Translated by H.P.Manning in Smith 1959, pp. 360-74.

LOOMIS, E. S. 1968: *The Pythagorean Proposition*. NCTM, Washington, D.C., 1968.

MADDY, P. 1988a: 'Believing the Axioms I'. *The Journal of Symbolic Logic*, vol. 53, no. 2, June 1988. pp: 481-509.

__________ 1988b: 'Believing the Axioms II'. *The Journal of Symbolic Logic*, vol. 53, no. 3, September 1988. pp: 736-764.

__________ 1990: *Realism in Mathematics*. Oxford: Clarendon Press, 1990.

__________ 1991: 'Philosophy of Mathematics: Prospects for the 1990s'. *Synthese*, 88, 1991, pp. 155-164.

__________ 1992: 'Indispensability and Practice'. *Journal of Philosophy*, vol. 89, no. 6, June 1992. pp: 275-289.

__________ 1997: *Naturalism in Mathematics*. Oxford: Clarendon Press, 1997.

MALAMENT, D. 1982: Review of Hartry Field: *Science Without Numbers* [Field 1980]. *Journal of Philosophy*, vol. 79, pp. 523-34, 1982.

MANCOSU, M. 1991: 'On the Status of Proofs by Contradiction in the Seventeenth Century'. *Synthese*, 88, 1991, 15-41.

MELIA, J. 1998: 'Field's Programme: Some Interference'. *Analysis*, vol. 58, no.2, 1998, pp.63-71, 1998.

MEYER, R. K AND SLANEY, J. K 1989: 'Abelian Logic (from A to Z)' in *Paraconsistent Logic - Essays on the Inconsistent*. Graham Priest, Richard Routley, Jean Norman (eds). 1989.

MILL, J.S. 1884: *A System of Logic*. London: Longmans, 1970.

MILLER, D. 1974a: 'On the Comparison of False Theories by their Bases'. *British Journal for the Philosophy of Science*, vol. 25, 1974, pp. 178-188.

_________ 1974b: 'Popper's Qualitative Theory of Verisimilitude'. *British Journal for the Philosophy of Science*, vol. 25, 1974, pp. 166-177.

MOORE, G.H. 1982: *Zermelo's Axiom of Choice*. New York: Springer, 1982.

_________ 1989: 'Towards a History of Cantor's Continuum Problem'. In D. Rowe and J. McCleary (eds.) *The History of Modern Mathematics. Vol. 1: Ideas and their reception*. Boston: Academic Press, pp. 79-121.

MORTENSEN, C. 1998: 'On the Possibility of Science Without Numbers'. *Australasian Journal of Philosophy*, vol. 76, no. 2, 1998, pp. 182-97.

NAGEL, E. 1979: 'Impossible Numbers'. In *Teleology Revisted*. Columbia University Press, 1979.

NERLICH, G. 1976: *The Shape of Space*. Cambridge University Press, 1976.

NEWMAN, J. (ed.) 1956: *The World of Mathematics*. Simon & Schuster, 1956.

PARSONS, C. 1977: 'What is the Iterative Conception of Set?'. In Benacerraf and Putnam 1983, pp. 503-29.

_________ 1990: 'The Structuralist View of Mathematical Objects'. *Synthese*, 84, 1990, pp. 202-346.

PEANO, G. 1889: *Arithmetices principia, nova methodo exposita*. Translated as 'The Principles of Arithmetic, Presented by a New Method' by J. Van Heijenoort in van Heijenoort 1967.

PENROSE, R. 1989: *The Emperor's New Mind*. Oxford University Press, 1989.

POINCARÉ, H. 1893: 'Sur la généralisation d'un théorème d'Euler relatif aux polyèdres'. *Comptes rendus des séances de l'Académie de Sciences*, 117, p. 144.

_________ 1894: 'Sur la nature du raisonnement mathématique'. *Revue de métaphysique et de morale*, vol. 2. Translated as 'On the Nature of Mathematical Reasoning' in Poincare 1907. Excerpted in Benacerraf and Putnam 1983.

_________ 1907: *Science and Hypothesis*. Translated by W. J. Greenstreet; reprinted, New York: Dover Press, 1952.

POLYA, G. 1954a: *Mathematics and Plausible Reasoning. Volume 1: Induction and Analogy in Mathematics*. Princeton University Press, 1954.

_________ 1954b: *Mathematics and Plausible Reasoning. Volume 2: Patterns of Plausible Inference*. Princeton University Press, 1954.

POPPER, K. 1959: *The Logic of Scientific Discovery*. London: Hutchinson, 1959.

_________ 1963: *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge, London, 1963.

PRIEST, G. 1987: *In Contradiction*. Dordrecht: Nijhoff, 1987.

PUTNAM, H. 1965: 'Craig's Theorem'. *Journal of Philosophy*, vol. 62, 1965. Reprinted in Putnam 1975a, pp. 228-236.

_________ 1971: *Philosophy of Logic*. New York: Harper; London: George Allen and Unwin, 1971. Reprinted in Putnam 1975a, Second Edition, pp. 323-357.

_________ 1973: 'Meaning and Reference'. *Journal of Philosophy*, vol. 70, pp. 699-711.

_________ 1974: 'The 'Corroboration' of Theories'. In P.A Schilpp (ed.) *The Philosophy of Karl Popper*. Open Court Press, 1974. Reprinted in I. Hacking (ed.) *Scientific Revolutions*. Oxford Readings in Philosophy Series. Oxford University Press, 1981, pp. 60-79 and Putnam 1975a, pp. 250-269.

_________ 1975a: *Mathematics, Matter and Method: Philosophical Papers Volume 1*. Cambridge University Press, 1975.

_________ 1975b: *Mind, Language and Reality: Philosophical Papers Volume 2*. Cambridge University Press, 1975.

_________ 1975c: 'What is Mathematical Truth?'. In Putnam 1975a, pp. 60-78.

_________ 1975d: 'Mathematics Without Foundations'. In Putnam 1975a, pp. 43-59. First published in *Journal of Philosophy*, vol. 64, 1967. Reprinted in Benacerraf and Putnam 1983, pp. 295-311.

_________ 1975e: 'An Examination of Grünbaum's Philosophy of Geometry'. In Putnam 1975a, pp. 93-129.

_________ 1975f: 'Explanation and Reference'. In Putnam 1975b.

_________ 1975g: 'The Meaning of 'Meaning''. In Putnam 1975b.

_________ 1975h: ''Degree of Confirmation' and Inductive Logic'. In Putnam 1975a, pp. 270-292. First published in *The Philosophy of Rudolf Carnap* (ed.) P.A.Schilpp. Open Court, 1963.

_________ 1975i: 'Probability and Confirmation'. In Putnam 1975a, pp. 293-304. First published in *The Voice of America, Forum Philosophy of Science*, vol. 10, 1963.

_________ 1994a: *Words and Life*. Harvard University Press, 1994.

_________ 1994b: 'Artificial Intelligence: Much Ado About Not Very Much'. In Putnam 1994a.

_________ 1994c: 'Rethinking Mathematical Necessity'. In Putnam 1994a.

QUINE, W. V. 1936: 'Truth By Convention'. In Quine 1966, Benacerraf and Putnam, 1983.

_________ 1937: 'New Foundations for Mathematical Logic'. *American Mathematical Monthly*, vol. 44. Reprinted in Quine 1980.

_________ 1940: *Mathematical Logic*. Norton, 1940; revised edition, Harvard University Press, 1951; reprinted, Harper Torchbooks, 1962.

_________ 1948: 'On What There Is'. *Review of Metaphysics*, vol. 2, 1948. Reprinted in Quine 1953.

_________ 1951: 'Two Dogmas of Empiricism'. *Philosophical Review*, vol. 60, 1951. Reprinted in Quine 1953, Hart 1996.

_________ 1953: *From a Logical Point of View: 9 Logico-Philosophical Essays*. Harvard University Press, 1953. Second Edition, Revised, Harvard University Press, 1980.

_________ 1957: 'The Scope and Language of Science'. *British Journal for the Philosophy of Science*, vol. 8, no. 29, 1957, pp. 1-17. Reprinted in Quine 1966.

_________ 1958: 'The Philosophical Significance of Modern Logic'. In *Philosophy in the Mid-Century: A Survey. Vol I. Logic and Philosophy of Science*. R. Kiblansky (ed.). Florence: La Nuova Italia, 1958

_________ 1962: 'Carnap and Logical Truth'. In Quine 1966, Benacerraf and Putnam 1983.

_________ 1966: *The Ways of Paradox and Other Essays*. New York: Random House, 1966 and London: Harvard University Press, 1976.

_________ 1969a: 'Epistemology Naturalized'. In Quine 1969b.

_________ 1969b: *Ontological Relativity and Other Essays*. Columbia University Press, 1969.

_________ 1975: 'Five Milestones of Empiricism'. In Quine 1981.

_________ 1980: *From a Logical Point of View*. Harvard University Press, 1980. Revised, Second Edition of Quine 1953.

_________ 1981: *Theories and Things*. Harvard University Press, 1981.

_________ 1987: *Quiddities: An Intermittently Philosophical Dictionary*. Harvard University Press, 1987.

_________ 1992: *Pursuit of Truth*. Harvard University Press, 1992.

RAWLS, J. 1971: *A Theory of Justice*. Oxford University Press, 1971.

READ, S. 1988: *Relevant Logic*. Blackwells, 1988.

_________ 1994. *Thinking About Logic*. Oxford University Press, 1994.

REICHENBACH, H. 1958: *The Philosophy of Space and Time*. New York, 1958.

RESNIK, M. 1981: 'Mathematics as a Science of Patterns: Ontology and Reference'. *Noûs*, 15, 1981, pp. 529-50.

_________ 1982: 'Mathematics as a Science of Patterns: Epistemology'. *Noûs*, 16, 1982, pp. 95-105.

_________ 1984: Review of Charles Parson's *Mathematics in Philosophy*. *Journal of Philosophy*, 81, pp. 783-94.

_________ 1992: 'Proof as a Source of Truth'. In Detlefsen (ed.) 1992.

RESNIK, M. AND KUSHNER, D. 1987: 'Explanation, Independence and Realism in Mathematics'. *British Journal for the Philosophy of Science*, 38, 1987, 141-158.

RICHSTEIN, J. 2000: 'Verifying Goldbach's conjecture up to 4×10^{14}.' *Mathematics of Computation*, July 2000. Also available at: http://www.mcs.dal.ca/~joerg.

RIEMANN, B. 1854: 'On the hypotheses which lie at the Foundations of Geometry'. Translated by H.S.White and reprinted in Smith (ed.) 1959, pp. 411-425.

_________ 1859: 'On the Number of Primes Less than a Given Magnitude'. Translated in Edwards 1974, pp. 299-305.

RUSSELL, B. 1903: *The Principles of Mathematics*. Cambridge University Press; 2nd ed., London: Allen and Unwin, 1937.

_________ 1905: 'On Denoting'. *Mind*. Vol. 14, 1905.

RUSSELL, B. AND WHITEHEAD, A. N. 1910-13: *Principia Mathematica*. 3 vols. Cambridge University Press; 2nd ed., 1925-7.

SACCHERI, G. 1733: *Euclid Freed From Every Flaw*. Excerpted in Smith 1959, pp. 351-9.

SALMON, W. 1989: *Four Decades of Scientific Explanation*, University of Minnesota Press, 1989.

SANDBORG, D. 1998: 'Mathematical Explanation and the Theory of Why-Questions'. *British Journal for the Philosophy of Science*, vol. 49, 1998, pp. 603-24.

SAWYER, W.W. 1964: *Vision in Elementary Mathematics*. Introducing Mathematics series, no.1. Penguin Books, 1964.

SCHARLAU, W. AND OPOLKA, H. 1985: *From Fermat to Minkowski. Lectures on the Theory of Numbers and its Historical Development*. New York: Springer-Verlag, 1985.

SCHILPP, P.A. (ed). 1974: *The Philosophy of Karl Popper*. 2 vols. Open Court Press, 1974.

SHAPIRO, S. 1983a: 'Conservativeness and Incompleteness'. *Journal of Philosophy*, 80, 1983, pp: 521-531.

_________ 1983b: 'Mathematics and Reality'. *Philosophy of Science*, 50, 1983, pp. 528-48.

_________ 1989: 'Logic, Ontology and Mathematical Practice'. *Synthese*, 79, 1989. pp. 13-50.

SILVERMAN, J. H. 1996: *A Friendly Introduction to Number Theory*. Prentice Hall, 1996.

SKOLEM, T. 1922: 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre'. Translated as 'Some Remarks on Axiomatized Set Theory' in van Heijenoort 1967, pp. 290-301.

SINGH, S. 1998: *Fermat's Last Theorem*. London: Fourth Estate, 1998.

SMITH, D. E. 1959: *A Source Book in Mathematics*. New York: Dover Press, 1959.

SOBER, E. 1993: 'Mathematics and Indispensability'. *Philosophical Review*, vol. 102, no. 1, 1993.

STEBBING, S. 1939: *Thinking to Some Purpose*. Penguin Books, 1939.

STEINER, M. 1973: 'Platonism and the Causal Theory of Knowledge'. *Journal of Philosophy*, 70, 1973.

_________ 1975: *Mathematical Knowledge*. Cornell University Press, 1975.

_________ 1978: 'Mathematical Explanation'. *Philosophical Studies*, vol. 34, 1978, pp. 135-151.

STEWART, I. 1987: *The Problems of Mathematics*. Oxford University Press, 1987.

STICH, S. AND NISBETT, R. 1980: 'Justification and the Psychology of Human Reasoning'. *Philosophy of Science*, 47, pp. 188-202.

SUPPES, P. 1972: *Axiomatic Set Theory*. New York: Dover Press, 1972.

TARSKI, A. 1944: 'The Semantic Conception of Truth and the Foundations of Semantics'. *Journal of Philosophy and Phenomenological Research*, 4, 1944, pp. 341-75. Reprinted in Feigl and Sellars 1949 and Linsky 1952.

_________ 1956: 'The Concept of Truth in Formalized Languages' in his *Logic, Semantics and Metamathematics*. Translated by J. H. Woodger, Oxford University Press, 1956.

TICHÝ, P. 1974: 'On Popper's Defintions of Verisimilitude'. *British Journal for the Philosophy of Science*, vol. 25, 1974, pp. 155-160.

_________ 1978: 'Verisimilitude Revisited' in *Synthese*, vol. 38, 1978, pp. 175-196.

VAN DE LUNE, J, TE RIELE, H.J.J AND WINTER, D.T 1986: 'On the zeros of the Riemann zeta function in the critical strip (IV)'. *Mathematics of Computation*, vol. 46, 1986, pp. 667-681.

VAN FRAASSEN, B. 1977: 'The Pragmatics of Explanation'. *American Philosophical Quarterly*, vol. 14, 1977, pp. 143-50.

_________ 1980: *The Scientific Image*. Oxford: Clarendon Press, 1980.

VAN HEIJENOORT, J. (ed.) 1967: *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press, 1967.

WAGSTAFF, S. 1978: 'The Irregular Primes to 125000'. *Mathematics of Computation*, vol. 32, 1978, pp. 583-91.

WANG, H. 1957: 'The Axiomatisation of Arithmetic'. *Journal of Symbolic Logic*, 22, 1957, pp. 145-57.

WILSON, M. 1992: 'Frege: The Royal Road From Geometry'. *Noûs*, vol. 26, no. 2, 1992, pp. 149-80. Reprinted in Demopoulos 1995.

WITTGENSTEIN, L. 1933: *Remarks on the Foundations of Mathematics*. Ed. G. H von Wright, R. Rhees and G. E. M Anscombe. Translated by G. E. M Anscombe. MIT Press, 1978.

__________ 1953: *Philosophical Investigations*. London: Macmillan, 1953.

__________ 1969: *On Certainty*. G. E. M Anscombe and G. H. von Wright (eds.). Oxford: Blackwell, 1969.

WRIGHT, C. 1983: *Frege's Conception of Numbers as Objects*. Scots Philosophical Monographs no. 2. Aberdeen University Press, 1983.

ZERMELO, E. 1904: 'Proof that every set can be well-ordered'. Translated by Stefan Bauer-Mengelberg, in van Heijenoort 1967, pp.139-141.

__________ 1908a: 'Investigations in the foundations of set theory I'. Translated by Stefan Bauer-Mengelberg, in van Heijenoort 1967, pp. 199-215.

__________ 1908b: 'A new proof of the possibility of a well-ordering'. Translated by Stefan Bauer-Mengelberg, in van Heijenoort 1967, pp. 183-98.