A DEFENCE OF MATHEMATICAL PLURALISM

E.B. Davies

16 March 2004

Abstract

We approach the philosophy of mathematics via an analysis of mathematics as it is practised. This leads us to a classification in terms of four concepts, which we define and illustrate with a variety of examples. We call these concepts background conventions, context, content, and intuition.

1   Introduction

Mathematics has been subjected to such intense scrutiny over many years that it now needs whole books to describe the variety of philosophical positions which can be taken towards its foundations. In spite of this there is no agreement even about whether one should adopt a broadly realist (Platonic) or anti-realist position towards its basic entities. Gödel's theorem is accepted as mathematically correct but its interpretation – whether it establishes that proof and truth are distinct – divides the philosophical community. Some think that the truth of Peano arithmetic is self-evident while others consider that we are forced to accept it by the nature of our intelligence.

In this paper we by-pass these problems temporarily, and describe mathematics as it is actually practised. We provide an analysis in terms of four concepts, background conventions, context, content, and intuition, and describe the relationship between these. We consider that our analysis provides a richer view of the subject than the usual approaches, and even that it demonstrates the irrelevance of some of the existing controversies about the subject.

We reiterate the comments of Lakatos [1976] that philosophers' concentration on foundations does not provide a true picture of the subject. Most mathematicians may not have much more understanding of the philosophy of their subject than crane drivers do of Newtonian mechanics, but, nevertheless, mathematics is, by

definition, what they do, not the tiny part of it which most philosophers find interesting. It is surely significant that Newton, and then Laplace, were able to calculate planetary motions with extraordinary accuracy long before the real numbers were rigorously defined. Their results needed no changes as a result of the foundational 'revolution' at the start of the twentieth century.

Focussing on the formal aspects of arithmetic or logic to the exclusion of everything else in mathematics is similar to regarding grammar as being the basis of spoken communications, as opposed to ideas. It is worth recalling that classical Greek geometry was highly sophisticated and preceded the formulation of Peano's axioms by over 2000 years. The Greek and later Arabic study of polynomial equations culminated in Cardan's complete solution of cubic equations using complex numbers in 1539. Even this preceded the first statement of the law of induction, by Maurolico in 1575. The contention that counting and the abstract law of induction are the most basic ingredients of mathematics does not stand historical scrutiny.

Before starting our description of mathematics, we need to address some matters of terminology. We will often avoid using the term 'realism', because it means various things to different people. Platonism is now often called realism, although it is quite different from scientific realism; it is also occasionally called idealism. With some justification Popper claimed that his World 3 objects are real, because they affect human behaviour and through them other material bodies. He emphasised that after their conception World 3 objects are much more than just ideas – they exist in their own right and have objective properties. He gave the example of Frege's set theory, which was, he wrote in Popper [1977, p.57], inconsistent *before* Russell discovered that fact. Bishop [1967, p.10] also calls his own position straightforward realism, but we will call it constructivism. We call our position contextualism, with the risk of creating new confusions, since this term is already used in the philosophical literature. Contextualism in mathematics should not be confused with scientific relativism. Science is the study of the external world, while mathematics is one of the ways we systematize our thoughts. It is hardly surprising that our ability to think carefully is helpful when we are trying to understand the world, but our thought processes and the external world are quite different. Philosophical beliefs about the two need not be related.

2   Definitions

It is tempting to start by discussing the difference between syntax and semantics in mathematics, but we defer that. We actually start by defining the concept of **context**. By this we will mean the assumptions, axioms and definitions relevant to a particular field of mathematics. These may often be divided into two groups, mathematical assumptions and logical rules, but if the subject itself is formal logic the two cannot be distinguished.

Let us consider constructive mathematics as developed in Bishop [1967]. In this field one accepts Peano arithmetic, but not the law of the excluded middle. Bishop allowed iterative constructions, for example of the real number system, but required that limiting procedures should be under quantitative control. He had definite philosophical reasons for proposing this system, which he regarded as more cogent than classical mathematics, but these will be discussed in Section 4. All we need say here is that constructive mathematics has different theorems from classical mathematics, and, even when the theorems appear similar, the proofs are often different and have greater constructive content. In constructive mathematics the intermediate value theorem cannot be proved. The notion of compactness must be formulated with much more care than in classical mathematics, but the constructive version is still very powerful. Constructive mathematics and classical mathematics inhabit different mathematical universes.

The **background conventions** of pure mathematics are the procedures by which mathematicians agree that the proof of a theorem is correct subject to the assumed context. If the context does not include the law of the excluded middle then they must check that it is, indeed, not used in the proof being examined. If the axiom of choice is not present in the context, it must again not be used. Similarly for the parallel axiom if one is doing projective geometry, and for Robinson-type infinitesimals is one is doing classical mathematics. When checking proofs for correctness mathematicians use informal logical arguments. Each line of a proof is determinately valid or invalid, and if each step is valid than the proof as a whole is valid. The fact that true-false arguments are used to check proofs is entirely separate from whether the particular context involves two-valued logic or allows a decision procedure.

It is entirely possible to check proofs using a computer if they are written out in a sufficiently formal manner. One then needs to be convinced that the program does indeed formalize the methods that mathematicians use to check the correctness of proofs, so human judgement is not eliminated, only pushed back a stage. In practice mathematicians agree about what constitute correct proofs within a context. Disagreements often arise from the fact that the steps of most proofs are very informal, and may need to be expanded in order to confirm their correctness. Human beings are very poor at checking formal proofs, but in practice this does not seem to be a serious impediment to consensus.

Our third concept is **content**. This means the theorems, both individually and collectively as a theory, which have been proved in a particular context. When mathematicians say that a theorem is true they mean that they have read and understood the proof, or that they trust others to have done so. When they say that it is false they mean that they know of a counterexample, know that assuming its truth leads to a contradiction, or are confident that within a short space of time they could provide one of these. In other circumstances they say that they do not know. In other words they talk in entirely epistemological terms. They may have private views about the ontology of their subject, but this is irrelevant to their

professional interactions. They may also believe that a proof of some theorem will eventually be found, but beliefs are not publishable *as mathematics*, although they are occasionally allowed in academic journals as concessions to the authors.

Our final concept is **intuition**. This provides guidance about what types of theorem are likely to be provable within a particular context. It is hard to communicate in writing because of its very personal nature, and little evidence of its key role is to be found in articles in journals. This is not the place to describe the wide variety of intuitions. They are frequently highly geometrical, but may be purely syntactical, as for example when seeing a pattern in the coefficients of

$$\frac{1}{1 - x + x^2} = 1 + x - x^3 - x^4 + x^6 + x^7 - ...$$

See Henley [1995], Lakatos [1976] and Polya [1954] for discussions of this and other examples. Intuitions may also be based upon seeing the possibility of useful analogies, as between the commutator formula

$$[A, BC] = [A, B]C + B[A, C]$$

for matrices and Leibniz's rule

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x).$$

The similarity is again purely algebraic.

Some mathematicians have much better intuitions than others, but most remember occasions on which their intuitions have let them down, i.e. led them to think that a theorem was true when it was actually false (or vice versa). Such occasions may be resolved in one of two ways. The mathematician may modify his intuition, often quite a painful process. Alternatively he may alter the context of his theory, by changing some of the axioms or definitions. There is a closed loop in which the discovery of new content can change the intuition, the new intuition can lead to alterations of the context, while the context controls the content. None of these dominates. Indeed pure mathematics might be said to consist of the discovery or invention of such closed loops. Their interest is judged on the basis of their novelty and technical non-triviality.

An example of this was the formal definition of the real number system by Weierstrass and others. To the dismay of the inventors it turned out that their definition led to the conclusion that there existed continuous functions which were nowhere differentiable. They regarded such pathologies as very distasteful, but decided to accept the consequences of their own definitions and modify their intuitions rather than the definition of the real number system (the context of their theory). A hundred years later their approach is so ingrained in the way we teach analysis that it is difficult to understand why they had this problem. Indeed the nineteenth century nightmare of nowhere differentiable functions became the twentieth century heaven of self-similar sets and fractal dimensions.

We next clarify our position with respect to the two main questions of Tennant [1997, p.159]. On the first (could there be recognition transcendent truths?) we can be agnostic. The job of mathematicians is to provide proofs of those things which can be proved, and to classify those things which can be classified. To the second (is truth determinately bivalent?) we give a positive answer with respect to the verification of proofs using the background conventions of mathematics, in the sense defined above. Whether or not truth is bivalent for statements within a particular theory depends upon the context. In classical mathematics it is and in constructive mathematics it is not.

We now turn to the promised discussion of syntax versus semantics. Clearly the background conventions used to check proofs should be called syntax, while intuition lies in the semantic category. If one looks only at a completed theory it is easy to conclude that mathematics is entirely syntax. One sees only axioms, definitions and proofs of theorems. These are almost never written down in a completely formal manner, but the belief is that they could be. If one identifies mathematics with the final product of mathematical endeavours, one might well call oneself a formalist in the sense of Hilbert. However, mathematics is a goal-directed activity, and the context is set by the action of human intuitions. The definitions may be adjusted several times while the theory is being constructed, so the appearance of the final product is misleading. The more experienced a mathematician, the greater the extent to which he is able to skim through a proof looking for the ideas, only returning to check the details once he is sure that he understands the overall strategy. Usually this aspect is the main element of conversations between mathematicians. Linear reading of a text is reserved for situations in which one has so little understanding of a new field that nothing else is possible. From the lofty perspective of the experienced mathematician, a theory is defined by its intuitions, and the formal aspects are mere details.

In the next two sections we discuss the weaknesses of classical mathematics and the strengths of constructive mathematics. We pass by the strengths of classical mathematics and the limitations of constructive mathematics, because these are much better known.

3   Problems with Classical Mathematics

The pluralist approach to mathematics described above contrasts with the assertion by Maddy [1997, p. 191] that mathematicians have decided in favour of classical mathematics. For a very small number of mathematicians this means ZFC, but for most it means anything which may be proved using Peano arithmetic, naive set theory, the axiom of choice, and a naive notion of truth which implies the law of the excluded middle.

Gödel epitomized this view. He described himself as a Platonist, and considered

that sets had an objective existence. His own theorems proved the deficiencies of formal mathematical reasoning rather than the failure of certain statements to have any truth value. For Gödel ZFC was a partial list of properties of an external Platonic reality. He was confident that the continuum hypothesis was either true or false, because it referred to entities which either existed or did not.

The shaky philosophical foundations of classical mathematics were clear from its inception. There was a vigorous debate early in the twentieth century over the acceptability of the axiom of choice. Even its inventor, Zermelo, eventually agreed that the most compelling reason to accept it was the fact that without it mathematicians could not prove large numbers of results which they needed (Maddy [1997, p. 56]). They accepted it for pragmatic reasons, hoping that a justification would later appear. Now we know that there can be no proof of the axiom, but it seems that many mathematicians regard it as being so obvious that it needs no serious discussion. Cohen [1966], the person who finally proved that the continuum hypothesis was independent of ZFC, disagreed.

> Historically, mathematics does not seem to enjoy tolerating undecidable propositions. It may elevate such a proposition to the status of an axiom, and through repeated exposure it may become quite widely accepted. This is more or less the case with the Axiom of Choice. I would characterize this tendency quite simply as opportunism. It is of course an impersonal and quite constructive opportunism. Nevertheless, the feeling that mathematics is a worthwhile and relevant activity should not completely erase in our minds an honest appreciation of the problems which beset us. (Cohen [1971])

The paradoxical consequences of accepting the 'truth' of classical mathematics have become clearer as time has passed. Very recently Friedman has discovered a series of straightforward finite theorems in mathematics which can only be proved on the assumption that certain very large cardinals exist. One can take several philosophical attitudes towards this mathematical result, but the author himself wrote that his

> findings raise the specific issue of what constitutes a valid mathematical proof and the general issue of objectivity in mathematics in a down to earth way. (Friedman [1988])

In our language he was questioning the context, not the content, of his results: there is little doubt that his proofs are valid in the limited sense that the steps follow as he claims under the hypotheses and rules of inference which he adopts. The issue is whether his theorems are so counter to intuition that they demand a revision of the context of the proofs. This could, of course, go in several directions, but Friedman expects that the use of new axioms in mathematics will steadily increase as time passes. Those who regard large cardinals as outside the framework of

acceptable mathematics may be persuaded by his results to adopt a non-realist, or non-bivalent, view of mathematical truth.

At the present time most mathematicians work within classical mathematics. However, the number who are prepared to consider alternatives is increasing, possibly because of the influence of computers on the development of mathematics. Questions of computational complexity have greater importance than they used to, and force one to concentrate on methods of proof rather than abstract truth. The heyday of abstract, classical mathematics was in the 1960s, but the current tendency is to concentrate more on the classification of mathematical entities such as finite simple groups and 3-manifolds.

We remark in passing that Markov [1958] has proved that the homeomorphism problem for closed 4-manifolds is algorithmically insoluble: it may be reduced to the word problem for groups and then to Turing's halting problem. This has caused mathematicians to switch their attention to sub-classes of 4-manifolds for which progress may be possible. Their response to Markov's result has been purely pragmatic, in line with my comment above that their job is proving theorems, not grieving over theorems which cannot be proved.

Our final example of the paradoxical consequences of accepting classical mathematics is taken from Ruelle [1993, p. 148], but originates from Gödel himself. It is worth repeating because it illustrates the issues involved in this debate particularly clearly. The story has to be told in the language of a classical Platonist.

Given a Turing machine and a natural number n, one can consider the finite set of programs of length up to n. When put into the machine each program terminates in a finite length of time or continues for ever. Let $T(n)$ denote the maximum running time of those programs which do terminate. If there existed an algorithm for computing the function $T(\cdot)$ then there would be a decision procedure for the Halting problem: given a program of length n let it run until time $T(n) + 1$ and, if it has not halted, it never will.

The question is whether the function $T(\cdot)$ should be regarded as existing. A Platonist will answer yes, because he considers that the meaning of the definition is clear. An constructivist will answer no because its existence makes no contact with anything which could be useful: we know, thanks to Gödel and Turing, that $T(\cdot)$ is not computable, so its existence has no significance for us. The existence of $T(\cdot)$ is an elementary consequence of the assumption that every meaningful mathematical statement (in this case that every program either halts or does not halt in a finite but unspecified length of time) has a determinate value independent of our being able to know it.

Those who insist that the function $T(\cdot)$ exists have to admit that it increases with unimaginable rapidity. If it were bounded above by $2^{\wedge}(2^{\wedge}(2^{\wedge}(2^n)))$ or any other computable function, then by using that function in place of $T(\cdot)$ above one would have produced a decision procedure for the Halting problem.

4   In Support of Constructive Mathematics

There are many theorems in mathematical analysis whose proof depends upon some form of the axiom of choice. One of these is the Hahn-Banach theorem. This is the statement that if $\mathcal{L}$ is a closed linear subspace of a Banach space $\mathcal{B}$ and ϕ is a bounded linear functional on $\mathcal{L}$, then ϕ may be extended without increase of norm to $\mathcal{B}$. The standard proof depends upon using Zorn's lemma to 'find' a maximal extension of ϕ with the required properties, and then using a further one-step extension procedure to prove that this maximal extension already has domain equal to $\mathcal{B}$.

If one assumes that $\mathcal{B}$ is separable, then there is an alternative and more constructive approach. One can apply the one-step construction iteratively, at each step including one more element from the countable dense set which is provided. After a countable number of steps the functional has been extended to a dense subspace of $\mathcal{B}$, and one can apply a completion procedure.

This proof is not incapable of being criticised, in that it supposes that a countable number of successive steps can be completed, but it is much more constructive than the general proof. Many mathematicians consider that the classical proof is better than the constructive one, on the grounds that it applies in greater generality. We next suggest that this extra power is to some degree spurious.

One of the major applications of the Hahn-Banach theorem is to prove that every Banach space $\mathcal{B}$ has a Banach dual space $\mathcal{B}^*$ with various properties. The precise structure of $\mathcal{B}^*$ is of great importance, and has been determined for a large number of Banach spaces. Almost all of the spaces $\mathcal{B}$ for which the dual space can be described explicitly are separable. Two of the most important non-separable Banach spaces are $L^\infty(X, \mathrm{d}x)$, the space of essentially bounded functions on a measure space, and $M(\Omega)$, the space of signed measures on a compact Hausdorff space Ω. In addition the entire theory of von Neumann algebras deals with non-separable Banach spaces. In all these cases the description of the dual space is extremely abstract, and involves discussions in set theory which leave most mathematicians cold (but not all; see Kaplan [1985]). In particular the Banach dual space of $l^\infty(\mathbf{Z})$ has a description in terms of the very abstract Stone-Čech compactification of the integers.

As our second example we consider the intermediate value theorem for continuous functions of a single real variable. The standard proof of this theorem depends upon using the 'fact' that every bounded set of real numbers has a least upper bound. Bishop [1967, p. 5] provides compelling evidence that the intermediate value theorem is not valid in a constructive context. In classical terms the value of the variable x for which $f(x) = c$ exists, but there is no *general* algorithm for determining it from the function.

The classically trained mathematician does not regard this as disturbing, because for him existence has a Platonic meaning independent of computability. Bishop

took a very different line. Bishop [1967] followed Kronecker and Poincaré in holding that integer arithmetic was forced on us by the nature of our intelligence, but also considered that the law of the excluded middle should not be used, . He argued that humans should not concern themselves with the parts of mathematics which are in God's domain: as finite beings we should study problems which we can solve using procedures which we can control. He performed the very valuable service of proving, by doing it, that mathematics was not fatally weakened by taking this stance. Previously most mathematicians, observing Brouwer's failed advocacy of the same idea, had believed that avoiding the law of the excluded middle imposed unacceptable penalties.

However, Bishop's philosophy was not well worked out, and one does not need to buy into it totally in order to benefit from what he achieved (see Billinge [2003]). We can think of the difference between the two systems as depending upon different interpretations of the quantifier $\exists$. Its interpretation is classified under the heading of intuition, and should be distinguished from the rules governing its use, which are a part of the context of a theory. In classical mathematics $\exists$ refers to Platonic existence, but intuitionists consider that it should only be used to refer to the possibility of constructing the relevant quantity. At the formal level the difference between the two positions is that in classical mathematics $\exists$ may be defined in terms of $\forall$: the expression $\exists x A$ is logically equivalent to $\neg(\forall a \neg A)$. In intuitionist mathematics $\exists$ is a new quantifier with stricter conditions for its application.

The fact that most pure mathematicians are Platonists is not a substantial argument in favour of that position. Very few of them have thought seriously about the issues. Most are simply adopting the simplest interpretation of what they are doing, because they are not really interested in philosophical questions, and want to spend their time proving theorems. Human beings are notoriously bad at understanding their thought processes, and routinely assume that their mental models of reality are 'obviously' right, in spite of the fact that others find quite different beliefs equally obvious.

If one considers that mathematics is the creation of the human mind, one does not need to accept or reject either of the above interpretations of $\exists$. In the appropriate context each may be more fruitful than the other. Constructive proofs may be more limited in scope and harder to complete, but they provide more information. Their absence indicates that quantitative solutions of certain problems are not possible without further conditions. Nor is there any scientific need to take a position on the 'correct' meaning of the quantifier $\exists$ in mathematics. One can simply remember which theorems depend on which axioms, so that one is prepared for whatever one may encounter in the future. Any other attitude indicates either laziness or quasi-religious commitment.

A possible argument in favour of the classical interpretation of $\exists$ is that it has a wider scope than the constructive interpretation, and the more inclusive definition is necessarily the better. The weakness of this position is best illustrated

by considering the following example, one of many I could have chosen, in which fruitfulness has been judged more important than generality.

Functional analysis and operator theory can be developed in the context of Hilbert spaces, Banach spaces or even locally convex topological vector spaces. Each of these involves a system of axioms which is more general than the previous one in the list. One observes that the theory of LCTVS's has stagnated since its heyday in the 1950s, because most applications can be treated more satisfactorily in the other, more concrete, frameworks. From the point of view of the richness of the theories the order is reversed: Hilbert space operator theory has the richest structure, partly because of its large group of symmetries (unitary operators).

One can hardly deny that the notion of fruitfulness has been of major importance in the survival of axiom systems. Fruitfulness involves value judgements, and shows that mathematicians are involved in the same type of human enterprise as the rest of the human race. The historical importance of fruitfulness in the development of mathematics has been explored in some depth by Maddy [1997]. Unfortunately she makes no reference, positively or negatively, to the work of Bishop, Nelson, Parikh or Robinson, and comes to the conclusion that 'history shows that such an opponent (of the axiom of choice) would be swimming against the tide' (Maddy [1997, pp. 56]). She also asserts that history shows that mathematicians have dismissed the constructivist movement (Maddy [1997, p.191]). Perhaps one might suggest that the sea of knowledge can have many tides, each capable of being enjoyed by different members of the same community.

5 Avoiding Unnecessary Philosophical Choices

Mathematical pluralism states that different types of mathematics can co-exist. Several mathematicians have made important contributions in two mutually inconsistent contexts. For example, Bishop was a pioneer in constructive analysis, but prior to that he had done important work in the theory of sup-norm algebras, a subject intermediate between functional and complex analysis. Nelson did important work on finite arithmetic, but he had previously had key insights in probability and quantum field theory. (I am not attempting to summarize their life's work.)

However, both Bishop and Nelson developed strong views about the 'philosophically correct context' for mathematics. In 'Predicative Arithmetic' Nelson wrote

> What is at issue here is not the familiar construct of formal mathematics, but a belief in the existence of ω (the set of natural numbers) prior to all mathematical constructions. What is the origin of this belief? The famous saying by Kronecker that God created the numbers, all the rest is the work of Man, presumably was not meant to be taken

seriously. Nowhere in the Book of Genesis do we find the passage: And God said, let there be numbers, and there were numbers; odd and even created he them, and he said unto them be fruitful and multiply; and he commanded them to keep the laws of induction. No, the belief in ω stems from the speculations of Greek philosophy on the existence of ideal entities or the speculations of German philosophy on a priori categories of thought. (Nelson [1986, p. 80])

In spite of the above there is now a new generation of mathematicians who are willing to accept the value of all of these variations from the classical scheme. One can identify three strands of opinion about this. The first is that classical mathematics is the final arbiter of truth, but constructive methods are sometimes pragmatically useful because they provide more information. In this view constructive mathematics is simply a convenient formalization of the constructive approach, and has no foundational significance.

The second view is that constructive mathematics is the real thing, but that it may sometimes be helpful to use classical mathematics to provide intuition about what might be proved within constructive mathematics. Since classical mathematics is an extension of constructive mathematics, any theorem proved in classical mathematics will (almost surely) not be disproved in constructive mathematics.

The pluralist attitude, which is the one proposed in this paper, is that one need not make such philosophical choices. One should simply accept each mathematical theory on its merits, and judge it according to the non-triviality and interest of the results proved within it. According to Schechter, this is increasingly the attitude which logicians take towards their subject.

> The "existence" of f – or of any mathematical object, even the number "3" – is purely formal. It does not have the same kind of solidity as your table and your chair; it merely exists in the mental universe of mathematics. Many different mathematical universes are possible. When we accept or reject the Axiom of Choice, we are specifying which universe we shall work in. Both possibilities are feasible... (Schechter [2004])

The first two groups of mathematicians may be identified as Platonists and constructivists respectively. One still has to ask whether the pluralist view is philosophically coherent. One might regard mathematicians who adopt it as anti-realists in practice, whatever they say, because their willingness to contemplate mathematics with the continuum hypothesis or axiom of choice one day and without them the next shows that they do not have any deep beliefs about the existence of the relevant entities.

It is also possible to argue that a pluralist is a realist in denial, because in the last resort he accepts that the correctness of proofs is judged using two-valued classical

logic. Each step is determinately valid or invalid, and the proof is valid if and only if each step is valid. Note that one cannot correctly replace the word 'validity' by 'truth' in the above sentence. One might say that it is true that the proof is valid if and only if it is indeed valid, but this is inconsequential, because it contains no new content. The notion of validity used when checking proofs is very basic, and certainly does not involve any commitment to the law of induction. It is arguable whether it even implies the ability to count, although it does depend upon the ability to recognize the order in which a string of symbols is placed on a line.

Both of the above criticisms of contextualism result from attempts to force it into frameworks which do not fit it. There is an analogy with competitive games. If someone plays bridge one evening and chess the next, he could justifiably assert that both are valid ways of engaging his intellect, and that he feels richer for appreciating the subtleties of both. They have many things in common – the fact that the players take turns, follow specified rules and try to beat their opponents – but the two games do not need to compete for supremacy.

Mathematics is not the same as chess, because the goal in one field is to prove theorems and in the other to beat one's opponent. As World 3 objects, the fact that theorems are human constructs does not rob them of objective validity (Popper [1977]). Every chess player agrees about what is a checkmate for one player, just as every mathematician agrees about what is a valid proof. These are objective facts within the appropriate contexts.

Mathematicians do not need to commit themselves about the 'right way' of developing their subject. Indeed, they serve the scientific community best by providing as wide a variety of different ways of developing the subject as possible, in the hope that some will prove useful in the future. Perhaps Peano arithmetic will one day be regarded as much less interesting than finite arithmetic by generations who are far more deeply involved in computational issues. Perhaps twentieth century discussions about the continuum hypothesis will one day be regarded as totally misguided, similar in nature to discussions about how many angels can dance on the head of a pin. Or possibly future generations will decide to accept it as an indispensable part of mathematics, although unprovable. This would not make it true, any more than the laws of of Scotland are true.

6   Varieties of Arithmetic

A huge number of investigations of mathematical realism have tried, without success, to reconcile two matters: people's feeling that counting, and hence arithmetic, are basic concepts which are not man-made; and the fact that there is no way of explaining how we can know of them if they exist independently of the human species. If our knowledge is obtained by applying reason, then the application of that reason would lead us to the same conclusions even if they did not exist.

We refer to Balaguer [1998] for a discussion of these issues and references to the literature.

This controversy refuses to die. In a recent article Redhead has revived the claims of Gödel, Lucas and Penrose that human beings have powers denied to computers (Redhead [2004]). He even claims that there is an elementary argument for this, which does not rely upon the use of Gödel's theorems. In the preface to his article he states that 'as a result of the work of Gödel and Tarski we know that truth does not equate with proof'. He also holds that arithmetic is a formalization of properties of the 'intended model', to which one can properly attach notions of truth and falsity. The notion of truth is a very difficult one, and both of his statements above are certainly controversial (Tennant [1997]).

Redhead's gloss on Gödel's incompleteness theorem is very close to the statement of the theorem given in Manin [1977, p. 81] – that the set of 'true formulas' is not equal to the set of 'deducible formulas'. However, if one adopts an intuitionist mind-set it is not difficult to find where the notion of truth is slipped into the book. Most accounts of Gödels' theorem follow his own paper in saying that within any formalization of arithmetic there are simple propositions which are not decidable from within the system.

Mendelson [1964, p.143,144] separates the statement of Gödel's incompleteness theorem from his discussion of true but unprovable results. He also refers to the 'standard model' of arithmetic, which he admits is regarded by some as too precarious a basis for consistency proofs (Mendelson [1964, p. 107]). If one regards Peano arithmetic as reached by a process of idealization from what we actually experience when counting, one will certainly agree with this.

There are several different axiomatizations of arithmetic. Peano arithmetic comes in at least three flavours. They use the same axioms, but differ in the logical arguments which may be used. First order Peano arithmetic allows only a limited range of predicates, second order Peano arithmetic allows far more, while Bishop's approach to the subject does not allow the use of the law of the excluded middle (Bishop [1967]). These are different systems, allowing different theorems, and one is not compelled to decide that one of them matches the 'true nature' of the counting numbers better than the others.

There are other variations in the foundations of logic or mathematics which are of interest, such as the so-called 'finite arithmetic' of Parikh [1971], developed by a substantial number of other mathematicians, including Carbone [2000], Nelson [1986] and Sazanov [1995]. This is a more radical departure from the usual axiom systems, but it is of value in certain contexts, for example in relation to computational complexity. In this system exponentiation is avoided and numbers such as $10^{10^{100}}$ are considered infeasible.

The are several versions of finite arithmetic, but the underlying intuition is easily conveyed using decimal arithmetic. This was *invented* by Hindu-Arabic mathe-

maticians, and gradually diffused into Europe between 1000 AD and 1500 AD. Let $\mathcal{D}_{100}$ denote the system of decimal arithmetic in which each number has no more than a hundred digits, and with the standard rules for addition and multiplication. (The number 100 may be increased of course.) Each decimal number is simply defined as a string of digits, and one makes no commitment to the existence of the set of all such numbers. Nor need one assume that 'in principle' one could count from 1 up to 10^{100} if one were sufficiently determined. It is known that the law of induction may be proved within $\mathcal{D}_{100}$ in a few pages, using only routine logic and the law of substitution ([2003a]). Peano arithmetic is obtained by abstracting, or idealising, the properties of this system, but by itself $\mathcal{D}_{100}$ goes far beyond the needs of scientists, who cannot measure anything with greater than 15 digit accuracy. In fact the only uses for numbers with 100 digits are in cryptography, in which the connection with counting is irrelevant.

Another variation on classical mathematics is Robinson's non-standard analysis, which has had stochastic applications via the so-called Loeb measure construction (see Cutland [1988], Robinson [1979]). Although a minority interest, it is genuine mathematics which goes against the grain of the classical tradition by reviving the notion of the infinitesimal, which everyone thought had been killed by Cauchy in the 1820s. History has a habit of playing tricks on those who think that the 'right way' of setting up the foundations of a subject can be settled once and for all, even in mathematics.

At the other extreme from non-standard analysis is (rounded, floating point) interval arithmetic. This has an ancient history, but is becoming slowly more important in scientific computing, particularly in connection with computer assisted proofs of theorems in analysis; see Markov and Okumura [1999], Moore [1979] and the Interval Computations Web-Site [2004] for further details and some of the many applications. Its basic entities are rather like real numbers, and may be written in the form

$$x = 1.24706^{304}_{296} \ e3$$

where one pre-assigns the number of significant digits allowed. The interpretation of this expression is as the interval $[\underline{x}, \overline{x}]$ where $\underline{x} = 1247.06296$ and $\overline{x} = 1247.06304$, but the definitions of the basic operations of arithmetic on the entities do not depend logically upon the intuition, nor upon any commitment to the existence of the real number system. To add two entities one adds the lower bounds and then rounds *down* to the prescribed number of digits, and also adds together the two upper bounds and rounds *up* to the prescribed number of digits. One identifies an integer n with the interval $[n, n]$ and writes $x \sim y$ if the two intervals overlap, i.e. if

$$\max\{\underline{x}, \underline{y}\} \le \min\{\overline{x}, \overline{y}\}.$$

One puts $x > 0$ if $\underline{x} > 0$, and $x < 0$ if $\overline{x} < 0$; if neither of these holds then $x \sim 0$. One might define π in the system by $\pi \sim 3.141592655^4_3$, without commitment to the existence of an 'exact' value. In interval arithmetic

$$(x - 1)^2 \sim x^2 - 2x + 1$$

but the two are not equal. One needs to choose the 'right' way of evaluating an expression to minimize the width of the interval produced.

A systematic development of interval arithmetic has been completed, and programming languages using it are readily available. There is very little written about it of a philosophical character, but it could regarded as an alternative way of developing mathematics which does not necessitate commitment to the existence of infinite procedures. It is even possible to imagine it replacing standard real analysis at some future time.

Apparently bizarre variations in logic can be surprisingly useful. An example of this is paraconsistent logic, in which contradictions of the form (A and $\neg A$) do not imply every other formula B (Stanford Encyclopedia of Philosophy [2004]). Logics of this type can be useful when designing automated information processing systems, in which a computer is presented with large amounts of data, not all of which is correct or up to date. This may be the type of logic which humans use in their everyday life, and this may in turn be why we are able to contemplate different and mutually inconsistent logical systems without mental breakdown.

Nevertheless, most people would like assurances that their theorems or beliefs are consistent. They accept Gödels' proof that the consistency of arithmetic cannot be proved, but some believe that it is nevertheless consistent, because of the 'fact' that it describes actual properties of the 'expected interpretation' or the 'intended model'. Referring to the 'intended model' begs the question. Its existence is simply asserted without evidence, and therefore places one outside the scientific enterprise. The phrase 'expected interpretation' is more psychological in tone, but one is forced to comment that expectations do not always correspond to reality.

It seems to us entirely possible that Peano arithmetic is actually inconsistent. We have no evidence of this, and do not claim that it is likely to be inconsistent, only that it is possible. To support this idea we consider an example from group theory. Consider the following list of axioms.

(1) G is the set of elements considered, and it is supposed that the elements obey the group axioms.

(2) G is supposed to be finite but not isomorphic to any of the known list of finite simple groups.

(3) G is supposed to be simple. In other words if N is a subset which has certain list of properties (those of a normal subgroup other than the trivial subgroup) then $N = G$.

These axioms can be compared to those of Peano arithmetic. The last is similar in form to the induction axiom (or axiom schema in first order logic) in that it refers to an unspecified set with certain properties, and concludes that it is equal to G (we assume that one can switch back and forth between subsets and predicates). Although G is assumed to be finite its size is not specified, so one cannot simply

enumerate all objects of the above type, however long the time given: the only way of understanding the axiom system is via proofs.

I wrote down the above axiom system because it has one rather interesting feature. It is inconsistent, but the proof of this is likely to be about 3000 pages long: it is still in the process of being written down twenty five years after it was 'discovered' (Solomon [1995]). Of course the group theorists working on this problem were not thinking in such negative terms. They were trying to classify all finite simple groups, but their conclusion can be written, if rather perversely, in the above way.

Another example of an inconsistent axiom scheme is obtained by adding to Peano arithmetic the statement

> There exist positive integers a, b, c, n all larger than 10^{100}
> such that $a^n + b^n = c^n$.

Once again the proof of inconsistency is horrendously difficult, and the inconsistency of this extended Peano arithmetic has no implications for any results proved in classical mathematics before 1990.

The fact that axiom schemes so similar to Peano arithmetic might require such a long proof of their inconsistency provides a reason why we cannot be absolutely sure of the consistency of Peano arithmetic itself. Perhaps the shortest proof of an inconsistency in Peano arithmetic is one hundred million pages long, and we will never discover it. If we were never led into a contradiction, would the inconsistency matter? We could continue to prove theorems and derive interesting interconnections between ideas without ever suspecting the awful truth.

Such a situation need not imply that our efforts were worthless. There are many examples in the past in which contradictions in axiom systems, once pointed out, have been rectified. The most famous was Frege's set theory, in which Russell found a paradox. Within twenty years the ZFC set theory removed these particular problems, although at some cost in terms of elegance. Interesting mathematics (certainly in the field of analysis) is remarkably tolerant of changes in the axiomatic framework, and can also often be rescued from technical errors, possibly after changing or increasing the number of assumptions.

7   The Quine-Putnam argument

We conclude with some brief comments about the well-known argument that the objectivity of mathematics as a whole is guaranteed by its applications to the physical sciences. Even if one accepts this, we have presented evidence in Davies [2003a, 2003b] that one cannot derive the correctness of classical mathematics or prove the continuity of space from present-day physics. Physicists know that all of the mathematical models which they have constructed so far have only approximate

validity, and adopt a very pragmatic and empirical attitude towards them. For most of them proving theorems is simply not important. The shape of the final 'theory of everything' is entirely open, and many scientists do not believe that such a theory exists.

It is possible to argue that group theory is imposed by us on the physical world for our convenience, as opposed to its being a feature of the world itself. We know, for example, that Euclidean geometry is not even approximately relevant in parts of relativistic space-time close to black holes, but near to us it is very useful, even though we no longer regard space and time as separable. A model constructed by the author shows that the predictions of quantum theory are consistent with space being discrete (Davies [2003b]); in this model the apparent, continuous, Euclidean symmetries of space are emergent properties in the large scale limit. The physicist R Newton put it the following way.

> Conceivably, the resulting predictions could be made without using the concepts and results of the theory of groups; at the very least, however, to arrive at them in such a way would take enormously more time and effort. It is not that Nature itself makes use of group theory; it is that we humans need this invaluable mental crutch to understand Nature. Mathematics is not embedded in the structure of reality, but we require the help of its power to penetrate and describe that reality. (Newton [1997, p. 141]).

There is now a vigorous mathematical theory of large cardinals, but many mathematicians instinctively distrust it, and most physicists have probably never heard of it. Gillies [[2000] goes so far as to declare that this theory makes no contact with science and that its truths cannot be defended from an empiricist point of view. Certainly it is hard to find anything in physics which would be affected by the truth or falsity of the continuum hypothesis.

The Banach-Tarski paradox of 1924 states that it is possible to divide a solid ball into six parts which may then be moved around rigidly and combined into two balls, each the same size as the original. This physically absurd result depends upon using the axiom of choice to produce parts which are not Lebesgue measurable sets. Most physicists would say that it results from taking the mathematical model too seriously. Non-measurable sets do not qualify as potentially material objects. In addition no set which has diameter smaller than the Planck length has any relevance to physics, because at that scale classical language is not the appropriate way to describe physics. The obsession of mathematicians with the continuum limit is simply not appropriate in physics, where it is accepted only because it makes the equations simpler to write down and solve.

References

[1998] Balaguer M: 'Platonism and anti-Platonism in Mathematics'. Oxford Univ. Pess, Oxford, 1998.

[2003] Billinge H: Did Bishop have a philosophy of mathematics? Phil. Math. 11 (2003) 176-194.

[1967] Bishop E: 'Foundations of constructive analysis', McGraw-Hill, 1967.

[2000] Carbone A: Cycling in proofs and feasibility. Trans. Amer. Math. Soc. 352 (2000) 2049-2075.

[1966] Cohen, P. J. 'Set Theory and the Continuum Hypothesis'. W. A. Benjamin, New York, 1966.

[1971] Cohen P J: Comments on the foundations of set theory. p 9-15 in 'Axiomatic Set Theory', Proc. Symp. Pure Math. Vol XIII, Part I. Amer. Math. Soc., Providence, RI, 1971.

[1988] Cutland N: 'Nonstandard Analysis and its Applications', LMS Student Texts, Camb. Univ. Press, Cambridge, 1988.

[2003a] Davies E B: Empiricism in Arithmetic and Analysis. Philosophia Mathematica (3) 11 (2003) 53-66.

[2003b] Davies E B: Quantum mechanics does not require the continuity of space. Stud. Hist. Phil. Mod. Phys. 34 (2003) 319-328.

[1988] Friedman H M: Finite functions and the necessary use of large cardinals. Ann. Math. 148 (1998), 803893.

[[2000] Gillies D A: An empiricist philosophy of mathematics and its implications for the history of mathematics. pp. 41-57 in Emily Grosholz and Herbert Breger (eds.) 'The Growth of Mathematical Knowledge', Synthese Library, Volume 289, Kluwer, 2000.

[1995] Henley D S: Syntax-directed discovery in mathematics. Erkenntnis 43 (1995) 241-259.

[2004] Interval Computations Web-Site. http://www.cs.utep.edu/interval-comp/main.html

[1985] Kaplan S: 'The Bidual of $C(X)$ I'. North-Holland, Amsterdam, 1985.

[1976] Lakatos I: 'Proofs and Refutations: The Logic of Mathematical Discovery'. Camb. Univ. Press, Cambridge 1976.

[1997] Maddy P: 'Naturalism in Mathematics'. Clarendon Press, Oxford, 1997.

[1977] Manin Yu I: 'A Course in Mathematical Logic'. Springer-Verlag, New York, 1977.

[1958] Markov A A: Insolubility of the problem of homeomorphy, p. 300-306 in Proc. Intern. Congress of Math. 1958, Camb. Univ. Press, Cambridge, 1958.

[1999] Markov S, Okumura K: The Contribution of T. Sunaga to Interval Analysis and Reliable Computing. pp. 167-188 in: T. Csendes (ed.), 'Developments in Reliable Computing', Kluwer, Dordrecht, 1999.

[1964] Mendelson E: 'Introduction to Mathematical Logic'. Van Nostrand Reinhold, New York, 1964.

[1979] Moore R E: 'Methods and Applications of Interval Analysis'. SIAM, Philadelphia, 1979.

[1986] Nelson E: 'Predicative Arithmetic'. Princeton Univ. Press, 1986.

[1997] Newton R: 'The Truth of Science'. Harvard Univ. Press, Cambridge, MA, 1997.

[1971] Parikh R: Existence and feasibility in arithmetic. J. Symb. Logic 36 (1971) 494-508.

[1954] Polya G: 'Mathematics and Plausible Reasoning, Vol. 1'. Oxford Univ. Press, Oxford, 1954.

[1977] Popper K R, Eccles J C: 'The Self and Its Brain'. Routledge, London, 1977.

[2004] Redhead M: Mathematics and the mind, Brit. J. Phil. Sci. to appear.

[1979] Robinson A: 'Selected papers of Abraham Robinson. Vol. II. Nonstandard analysis and philosophy'. New Haven, Conn., 1979.

[1993] Ruelle D: Chance and Chaos'. Penguin, London, 1993.

[1995] Sazanov V Y: On feasible numbers, pp. 30-51 in: Leivant, D. ed., 'Logic and Computational Complexity', Lecture Notes Comp. Sci. Vol. 960, Springer, 1995.

[2004] Schechter E: 'Home Page for the Axiom of Choice'. http://www.math.vanderbilt.edu/~schectex/ccc/choice.html

[1995] Solomon R: On finite simple groups and their classification. Notices Amer. Math. Soc. 42 (1995) 231-239.

[1997] Tennant N: 'The Taming of the True', Oxford Univ. Press, Oxford, 1997.

[2004] Zalta E N (ed.): 'Stanford Encyclopedia of Philosophy'.
   http://plato.stanford.edu/entries/logic-paraconsistent/

Department of Mathematics,
King's College,
Strand,
London,
WC2R 2LS,
England.

E.Brian.Davies@kcl.ac.uk

http://www.mth.kcl.ac.uk/staff/eb_davies.html