

The mathematician as a formalist

H. G. Dales

1 Introduction

The existence of this meeting bears testimony to the anodyne remark that there is a continuing debate about what it means to say of a statement in mathematics that it is ‘true’. This debate began at least 2500 years ago, and will presumably continue at least well into the next millennium; it would be implausible and perhaps presumptuous to suppose that even the union of the talented and distinguished speakers that have been assembled here in Mussomeli will approach any solution to the problem, or even arrive at a consensus of what a solution would amount to.

In the end, it falls to the philosophers, with their professional expertise and training, to carry forward the debate and to move us to a fuller understanding of this subtle and elusive matter. Indeed, we are hearing at this meeting a variety of contributions to the debate from different philosophical points of view; also, there is a good number of recent published contributions to the debate (see (Maddy 1990), for example).

What then is the rôle of the mathematician in this debate? Some mathematicians take the view that, since they are doing mathematics, they certainly know what they are about—that ‘true mathematics’ is *ipso facto* what mathematicians are doing, and that philosophers have only the relatively minor rôle of clarifying what mathematicians know they are doing, mainly for the benefit of those unfortunate people who are not mathematicians. This must be an arrogant and mistaken view; there is no reason to suppose that mathematicians have an innate understanding of the philosophical foundations of their subject, or even any coherent and well-thought-out view of what exactly they are engaged in. Even those mathematicians who do believe they have such a coherent view of their subject may well find that, when this view is exposed to the scrutiny of a philosopher, the coherence

is illusory and that they must squirm as the inadequacies of their thoughts become apparent.

Thus we cannot expect mathematicians to resolve the problems of ‘truth in mathematics’ for the philosophers. But this does not mean that the thoughts and practices of mathematicians are irrelevant to the philosophers who are grappling with the problem: philosophers of mathematics must come to terms with mathematics as it is practised today, and they should not be content to base their theories on, say, the very different style of mathematics that was extant in the early years of this century. Modern philosophers of mathematics must understand the nature and content of modern mathematics, and have at least a nodding acquaintance with the working assumptions of modern mathematicians, elucidating the ‘best practices’, even if only to criticize the inadequacies of this *Weltanschauung*. (The role of ‘practitioners’ is discussed in (Maddy 1998c).)

There are two different aspects of modern mathematics that philosophers must take particular account of.

The first of these is the collection of specific theorems that mathematicians have proved within their subject. Even an apparently technical remark, such as Tychonoff’s theorem that an arbitrary product of compact topological spaces is compact, has philosophical significance. More obviously there are many great theorems of this century that have profound philosophical implications and that must be taken into account by modern philosophers; I am thinking, of course, of Gödel’s theorems, and the now classic proofs of the independence of the Axiom of Choice (AC) from ZF and of the Continuum Hypothesis (CH) from ZFC. However, there are now many recent and very significant specific results that are probably not well known to philosophers of which account must be taken; some of these, involving, for example, ‘large cardinal theory’ are also discussed in (Maddy 1998c).

The second aspect that surely must be taken into account is the style of presenting mathematics that is the orthodoxy of the present day, for this style should represent the (perhaps implicit) thoughts of the community of mathematicians about the fundamental nature of their subject; our perceptions may be naïve or mistaken, but philosophers should know what these perceptions are and why the underlying view is attractive to mathematicians, before criticizing and trying to lead us into a different direction.

My purpose here is to describe this modern orthodoxy and to explain how it has arisen and how it applies.

Thus I make a weak claim that the views I am expressing are those of a ‘normal’ working mathematician; it is true that I have made no survey of the views of these mathematicians, and I am well aware that many different views would be expressed by others—including other speakers at this meeting—and so probably I can only say that I am expressing my own views; I present myself as some sort of specimen. I should also explain how I am using the term ‘mathematician’. In this talk, I shall *exclude* from the set of mathematicians what is really the subset of mathematical logicians and set theorists¹ and deal with the complementary set. Also I am thinking particularly of practices within analysis and algebra; number theory may have a special rôle because of the perception that questions about natural numbers are particularly fundamental, and geometry and mathematical physics have a special connection with our philosophy of the physical universe.

2 Philosophical possibilities

First let me very briefly explain, perhaps caricature, some possible philosophical views, as seen by my mathematician. I apologize if these simplistic formulations give offence to adherents of particular doctrines.

A *realist*, or platonist, believes that the set-theoretic universe has an existence outside of ourselves, and hence that statements about this universe (such as the Continuum Hypothesis) are either true or false; that the axioms of ZFC are merely obviously true principles about sets that capture some of the total truth about real sets; that it is interesting that it is now known that we can neither prove nor refute CH from these axioms, but that this is not significant in deciding whether or not CH is true in the platonic universe. Thus it is possible for the realist to ask ‘Is the Axiom of Constructability ($V = L$) true?’ or ‘Do inaccessible cardinals exist?’, and to believe that these are meaningful questions. (I am probably here describing the *Simple Realism* of (Maddy 1998c); this sounds a little better than ‘naïve realism’.)

My main problem with this simple realism is that I cannot see how we could possibly decide on this basis whether or not CH is a true statement about the universe of sets: we can collect evidence, and discuss what would amount to evidence (cf. (Maddy 1998c), (Martin 1998)), but to *know* the truth of CH seems quite inaccessible to us.

The move to *formalism* at the beginning of this century came, I presume, in response to the emergence of paradoxes about the very notion of

set;² to avoid these paradoxes we must be more careful about the notion of ‘a property defining a set’ and our use of language. The solution, proposed by Frænkel and Skolem, consists in eliminating everyday language from mathematical statements, and replacing it by formal languages, hence ‘formalism’.

For a formalist, mathematics is the science of rigorous proof: we start from axioms chosen in some way; we hope that the axioms are not inconsistent; and we deduce what we can from the axioms by using a logical system that we have precisely delineated (probably first-order logic³) and by working in a formal language. The interpretation given to the axioms is irrelevant; we are concerned only with the validity of the deductions from them. Results proved in this way from the axioms are called ‘theorems’; incautiously, mathematicians tend to say that the theorems are ‘true’, but in fact the statements have no content, for they are not about anything, and ‘true’ is merely a brief way of saying that the theorems are what can be deduced from the axioms. A problem for the formalist is why we choose one set of axioms rather than another (we do quickly discard systems of axioms known to be inconsistent, but you will know that proofs that systems are consistent are not obtainable in the cases that interest us). Thus mathematics is seen, not as a science, but as a language; in Russell’s harsh phrase, it is a subject in which practitioners do not know what they are talking about and do not know whether or not what they are saying is true; in Dieudonné’s words ‘Mathematics is just a combination of meaningless symbols’. Thus I seem so far to be a *Glib Formalist* in the sense of Maddy (1998c); to quote Maddy, such persons hold that ‘all consistent theories are on a par, mathematically speaking, that the only justification an axiom requires is evidence for its consistency, that the choice between various axioms, between various theories of sets, is guided not by rational principles, but by aesthetic or psychological or sociological influences’. I do believe that all consistent theories have the same mathematical status, but I go beyond this in claiming that the choice among completing theories is not irrational; this is moving towards the *Subtle Formalism* of (Maddy 1998c).

The doctrine of *naturalism* is described in the previous chapter (Maddy 1998c); it will not be discussed further here.

It is not my present purpose (or within my capability) to engage in a serious philosophical defence of formalism in this talk; my intent is to describe how (some) mathematicians act as formalists when they are writing down their mathematics. Others can attempt to explain whether or not

these mathematicians are standing on firm philosophical ground.

There is a distinction for the formalist between the formal theory and the metatheory. For example, within ZFC, a (*formal*) *theorem* is a sentence in the formally prescribed language provable from the axioms of ZFC; a *metatheorem* is a statement about what can be proved in the formal theory.⁴ It seems that formalists are constructivists in respect of what can be proved in the metatheory.⁵

I shall speak only of the debate between realists and formalists. There are of course many nuanced versions of realism and formalism, and several other important philosophies of mathematics; some are expounded in other talks at this meeting. I will not discuss these here, save to say that I am not aware of any significantly large schools of working mathematicians who have adopted their tenets. For example, *finitism* has a certain appeal, but this point of view seems to discard much of modern mathematics. The case for *constructivism* is cogently presented by Bridges in this volume (Bridges 1998); clearly there is much beautiful mathematics here that has a wide appeal—for example, the constructive version of Picard’s theorem, described by Bridges, has been much appreciated—but it seems that only a small group of mathematicians has actively embraced the philosophical tenets of this doctrine and incorporated them into their own work.

3 Attitudes of mathematicians

The first remark must surely be that most mathematicians are, at best, rather indifferent to the debate between realists and formalists, and a good number is totally indifferent, or even antagonistic, to the existence of such philosophical musings. The extreme case is that of applied mathematicians and physicists, who, as Effros remarks in his lecture (Effros 1998), whilst valuing our language, often have little patience even for our insistence on rigour in proofs, and so these people are scarcely going to concern themselves with the difference between formalism and realism. But this is also so of (pure) mathematicians in my sense: a natural question for gossip in bars is ‘Is the cohomology theory of a von Neumann algebra necessarily zero?’, rather than ‘What does it mean to say that it is true that the cohomology is zero?’. But I will indicate below that questions of foundations can come and disturb even ‘normal’ mathematicians.

The second remark is one made several times before by other people:

mathematicians are ambivalent between realism and formalism. For example, I quote from Davis and Hersh (1981, p. 320):

... the typical working mathematician is a [realist] on weekdays and a formalist on Sundays. That is, when he is doing mathematics he is convinced that he is dealing with an objective reality whose properties he is attempting to determine. But then, when challenged to give a philosophical account of this reality, he finds it easiest to pretend that he does not believe in it after all.

Let me continue with a quotation from Yiannis Moschovakis (1980, p. 605):

Nevertheless, most attempts to turn these strong [realist] feelings into a coherent foundation of mathematics invariably lead to vague discussions of 'existence of abstract notions' which are quite repugnant to a mathematician. Contrast this with the relative ease with which formalism can be explained in a precise, elegant and self-consistent manner and you will have the main reason why most mathematicians claim to be formalists (when pressed) while they spend their working hours behaving as if they were completely unabashed realists.

The above two comments are certainly true at one level. However, I would change the emphasis from that in the first quotation. It seems to me that most mathematicians really are formalists for all the days of the week. It is of course very useful when seeking proofs within the formal system to have a 'realistic picture' in one's mind, and so it is temporarily convenient, during the week, to be a realist, but it is the realism that the mathematician does not really believe in. A proof is that which can be achieved within the formal situation, and not that which can be pictured in the image; even though one can become morally convinced of the validity of a general deduction by feelings that arise from consideration of the mental picture, the rôle of the mental construct is only psychological, and cannot convince in the written account that must eventually be produced if the insight is to find its place in the corpus of accepted mathematics, and not just be a private revelation. (This view contrasts with that of Jones in this volume (Jones 1998); it may very well be more applicable in areas of abstract analysis and algebra, which are my natural home, than in such geometric subjects as knot theory.)

I think that the success of the major mathematicians in resolving problems and advancing the subject owes much to their ability to formulate in their mind an appropriate image of the abstract problem: it must be sufficiently subtle and complicated to capture the essential features of the question at issue, yet remain sufficiently simple to allow our limited minds absolutely and fully to explore, in quiet contemplation, all aspects of this image until we understand it sufficiently to begin the attempt to transfer this understanding to a written account of the general, abstract situation. On the other side, I know that graduate students and all mathematicians sometimes falter because their intuitive, realistic image does not capture all relevant aspects of the question.

Thus my view is that we are genuine, believing formalists who temporarily act as realists for reasons of expediency in solving problems.

4 The style of formalists

I said earlier that philosophers should seek to understand the XXth century style of presenting mathematics; this has basically settled down since around 1930 to be the formalist's style. (As I have said, there are several penetrating critiques of this orthodoxy.)

The first remark is that formalists practically never use a truly formal language in their writings (and may not know how to do this, even under pressure); they formulate their theorems in the naïve language of set theory developed in the XIXth century by Dedekind and Cantor. But they are confident that, if their results had to be formalized, this could be done; and doubtless they are correct in this.

How then does a formalist choose his axioms and definitions? The choice of the axioms for set theory has been extensively discussed elsewhere, not least in other talks at this conference, and so I will draw my examples from other areas. Nevertheless it is clear that the fundamental axioms that underly the mathematics that I am talking about are the Zermelo–Fränkel axioms of set theory ZF, almost always taken with the Axiom of Choice AC to form the system ZFC; these axioms are listed by Woodin (1998). It could be said that the ‘axioms’ that I am presenting are merely abbreviations for concepts that arise in the theory ZFC: my point is to show examples of collections of axioms that mathematicians have chosen to delineate, and to

try to indicate why these particular collections of ‘meaningless symbols’ are so honoured.

The first example is that of a *group*. The systematic study of group theory dates from the early part of the XIXth century; it took a long time for the precise, abstract concept to be formulated. The formal definition now stands as follows.

Definition 4.1 *A group is a triple (G, \cdot, e) such that:*

- (i) *G is a non-empty set;*
- (ii) *$\cdot : G \times G \rightarrow G$ is a binary operation such that $r \cdot (s \cdot t) = (r \cdot s) \cdot t$ for all r, s, t in G ;*
- (iii) *e is an element of G such that $r \cdot e = e \cdot r = r$ for all r in G ;*
- (iv) *for each r in G , there exists s in G such that $r \cdot s = s \cdot r = e$.*

Certainly even this elementary definition uses words that need a prior definition. In particular, the definition of a group presupposes the definition of a set, and all that this implies.

It follows easily that the element e (called the *identity* of G) is uniquely specified by condition (iii) and that, for each $r \in G$, s is uniquely specified in condition (iv). These facts are very easily proved from the above axioms; the point is that in a careful exposition of group theory, they *must* be proved. Here is the proof that e is uniquely specified. Indeed suppose that e_1 and e_2 are elements of G that both satisfy the axiom (iii). By using two different equalities contained within (iii), we see that $e_1 \cdot e_2 = e_1$ and that $e_1 \cdot e_2 = e_2$, and so, by a more basic axiom, $e_1 = e_2$.

The notion of a group arose from the idea of permutations of a fixed (finite or infinite) set, the group operation being composition of permutations; these ideas, which are concerned with what we now call groups of permutations, arose in the early years of the XIXth century—Cauchy played a significant rôle—after much experimentation with specific results on the roots of polynomials in one variable, and, in particular, after the attempt ‘to solve the general polynomial of the fifth degree by radicals’.⁶ Of course examples of groups are ubiquitous in our mathematical world.

It is of fundamental importance to know when two groups are the same, or are isomorphic.

Definition 4.2 Two groups (G, \cdot, e_G) and (H, \times, e_H) are isomorphic if there is a bijection $\theta : G \rightarrow H$ such that $\theta(r \cdot s) = \theta(r) \times \theta(s)$ for each r and s in G .

It is this notion of isomorphism that underlies the great transformation in ideas of the XIXth century: we moved from the concept of mathematical *objects* (natural numbers for arithmetic, single equations for algebra, space and figures for geometry, specific functions in analysis) to that of *relations* between objects, epitomized by the notion of isomorphism. It is remarkable that apparently no one before 1850 noticed that the sets of real and complex numbers form a group with respect to addition, that the set of invertible $(n \times n)$ -matrices over \mathbb{C} forms a group with respect to composition, etc., and so the relations understood for hundreds of years in one context could easily spread to new situations.

My second example is that of a Hilbert space, arising from around 1920. I give the definition in a briefer form that begs the earlier definition of some terms.

Definition 4.3 A Hilbert space (over \mathbb{C}) is a linear space H over \mathbb{C} together with a complex inner product, mapping the pair (x, y) in $H \times H$ to the complex number $\langle x, y \rangle$ in \mathbb{C} , such that:

- (i) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ ($\alpha, \beta \in \mathbb{C}$, $x, y, z \in H$);
- (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ ($x, y \in H$);
- (iii) $\langle x, x \rangle \geq 0$ ($x \in H$);
- (iv) $\langle x, x \rangle = 0$ only when $x = 0$.

Further, the space H must be complete with respect to the associated norm defined by

$$\|x\| = \langle x, x \rangle^{1/2} \quad (x \in H).$$

Even for such an elementary object, the formal definition is a little complicated. Again, two Hilbert spaces H and K (we suppress the notation for the additional structure) are the ‘same’ if all the structures of H are indistinguishable from those of K : H and K are *isomorphic* if there is a

bijection T from H onto K such that T is a linear map over \mathbb{C} and T preserves the inner product in the sense that $\langle Tx, Ty \rangle = \langle x, y \rangle$ for each x and y in H .

The third example is that of a Banach algebra, first defined around 1940. I now give the definition in a decently terse form.

Definition 4.4 *A normed algebra $(A, +, \cdot, \|\cdot\|)$ is a structure such that:*

- (i) $(A, +, \|\cdot\|)$ is a normed space;
- (ii) $(A, +, \cdot)$ is a complex algebra;
- (iii) $\|ab\| \leq \|a\| \|b\| \quad (a, b \in A)$.

The structure is a Banach algebra if the normed space $(A, \|\cdot\|)$ is complete.

For example, let Ω be a compact space, and let $C(\Omega)$ denote the family of all continuous, complex-valued functions on Ω . Then $(C(\Omega), +, \cdot)$ is an algebra with respect to the obvious pointwise operations, and $(C(\Omega), +, \cdot, \|\cdot\|_\Omega)$ is a Banach algebra, where the *uniform norm* $\|\cdot\|_\Omega$ is defined by

$$\|f\|_\Omega = \sup\{|f(x)| : x \in \Omega\} \quad (f \in C(\Omega)).$$

When are two Banach algebras the ‘same’? There are now two variants of the basic definition: two Banach algebras A and B are *isomorphic* (respectively, *isometrically isomorphic*) if there is a bijection $\theta : A \rightarrow B$ such that

$$\theta : (A, +, \cdot) \rightarrow (B, +, \cdot)$$

is an algebra homomorphism and such that $\theta : (A, \|\cdot\|) \rightarrow (B, \|\cdot\|)$ is a continuous (respectively, an isometric) map.

The point of giving these definitions is to stress the fundamental view that a group, a Hilbert space, a Banach algebra is exactly what is specified by the definitions; they are no more and no less than this. Theorems about groups, Hilbert spaces, and Banach algebras are those results that can be deduced from the axiomatic definitions by the formal procedures that we allow; I do not see that we have any independent knowledge about these objects other than what can be proved in this way.

The definitions do come with an associated definition of when two objects are ‘the same’. In a sense it is unnecessary to state these additional

definitions because they can be subsumed under the diktat: ‘two structures in a category are isomorphic if there is a bijection between the underlying sets that preserves all the structures’.⁷ But note that we may have two somewhat different ‘isomorphic theories’: for example, in the case of Banach algebras we may chose to preserve the topological structure and work with isomorphic Banach algebras or to also preserve the geometric structure and work with isometrically isomorphic Banach algebras. If two structures are isomorphic, there is nothing that we can prove about the one that cannot be proved about the other.

These axiomatically defined objects are only useful and understandable if there are natural examples of the concepts.

There are two natural examples of a Hilbert space. First, let H_1 consist of the set of sequences $\alpha = (\alpha_n : n \in \mathbb{N})$ of complex numbers such that the sum

$$\sum_{n=1}^{\infty} |\alpha_n|^2$$

is convergent. Then H_1 is a Hilbert space with respect to coordinatewise linear space operations and the inner product defined by

$$\langle (\alpha_n), (\beta_n) \rangle = \sum_{n=1}^{\infty} \alpha_n \bar{\beta}_n \quad ((\alpha_n), (\beta_n) \in H_1).$$

Second, let H_2 consist of the family of Lebesgue measurable functions f on the closed unit interval $\mathbb{I} = [0, 1]$ such that

$$\int_0^1 |f(t)|^2 dt$$

is finite. Then H_2 is a Hilbert space with respect to the pointwise linear space operations and the inner product defined by

$$\langle f, g \rangle = \int_0^1 f(t) \overline{g(t)} dt \quad (f, g \in H_2).$$

(There is a certain subtlety about the second example, in that, strictly H_2 is the space of equivalence classes of functions, where $f \sim g$ if and only if

$$\int_0^1 |f(t) - g(t)| dt = 0;$$

so there is a certain ‘unreality’ about specifying an element of H_2 as a function.) For the formalist, H_1 and H_2 are the *same* Hilbert space because (this is not quite obvious, but not very deep) the two spaces are isomorphic. But looked at through other eyes, H_1 and H_2 are clearly very different. Does the realist have a concept of a Hilbert space? If so, which of my two examples is closer to the ‘real, platonic’ Hilbert space—or is it just the axioms which capture the essence of the ‘real, platonic’ Hilbert space? In the latter case, the difference of a realist from a formalist seems to evaporate. I see no reason why either example should be preferred to the other.

Again let me remark that two Banach algebras $C(\Omega_1)$ and $C(\Omega_2)$ are isomorphic if and only if the two compact spaces Ω_1 and Ω_2 are homeomorphic, but that, regarded as Banach spaces, $C(\Omega_1)$ and $C(\Omega_2)$ are isomorphic whenever Ω_1 and Ω_2 are both compact, uncountable metric spaces; the concept of ‘being the same’ depends on the structures that we take account of.

Let me detour briefly to give an example of a deduction from the axioms that gives pleasure to a mathematician. Look again at Definition 4.4. A Banach algebra has two different structures, in that it is both a Banach space and an algebra, and the two structures are related by the apparently weak condition 4.4(iii), which essentially asserts that the algebraic operation of taking the product is continuous with respect to the topology defined by the Banach space structure. However, it is a deep and beautiful result—it took about 25 years to evolve—that, under a simple algebraic condition, the Banach space structure is uniquely determined by the algebra structure; an algebra isomorphism is necessarily an isomorphism of Banach algebras.⁸

One could argue that the realist has no *a priori* concept of a Hilbert space or of a Banach algebra, and so they have no pressing necessity to pronounce on the intrinsic nature of these concepts of mathematicians. But surely the realist does have a concept of that fundamental construct, the real line \mathbb{R} ? What is the formalist’s real line? This depends on the aspects of the real line’s structure on which one wants to concentrate. For example my preferred definition is the following.

Definition 4.5 *A field is a structure $(K, +, \cdot, 0, 1)$ such that:*

- (i) $(K, +, 0)$ is an abelian group;
- (ii) $(K \setminus \{0\}, \cdot, 1)$ is an abelian group;

(iii) *the distributive laws hold.*

An ordered field is a structure $(K, +, \cdot, 0, 1, \leq)$ such that:

(i) *(K, \leq) is a totally ordered set;*

(ii) *$(K, +, \cdot, 0, 1)$ is a field;*

(iii) *$a + c \leq b + c$ whenever $a, b, c \in K$ with $a \leq b$;*

(iv) *$ab \geq 0$ whenever $a, b \in K$ with $a, b \geq 0$.*

An ordered field $(K, +, \cdot, 0, 1, \leq)$ is (Dedekind) complete if each non-empty subset of K which is bounded above has a supremum.

This is the standard definition of a complete ordered field that is offered (or, at least, used to be offered), with some preparation, to first-year students. We have an immediate definition, as in the above diktat, of when two ordered fields are isomorphic. But now we have a clear theorem: *any two complete ordered fields are isomorphic to each other.* Thus my view is that any two complete ordered fields are the same, and so there is just one such field; this field has the properties that one would wish the real line \mathbb{R} to have; and so, by its very definition, \mathbb{R} is exactly ‘the’ complete ordered field; the properties of \mathbb{R} are the theorems about the structure ‘complete ordered field’. Note immediately that these properties do not include a resolution of the question whether or not there is an uncountable subset of \mathbb{R} which is not equipotent to the whole of \mathbb{R} ; the notion of isomorphism that flows from the structure I have chosen to call that of \mathbb{R} is not refined enough to carry this extra information. I will describe shortly how I believe we should proceed in deciding this matter.

It will be said that there are other definitions of the real line that capture different properties, perhaps that others consider to be more important. This is indeed the case. My view is exactly this: the idea of the real line is the inspiration of many different topics within mathematics, and can be captured by different sets of axioms; that when one talks of \mathbb{R} as a complete ordered field, its properties are just the theorems about such fields; but when one characterizes \mathbb{R} by different axioms, one obtains a different collection of properties.

Presumably the realist’s real line has the union of the properties that have been formulated, and others not yet, or perhaps never to be, known.

One of these properties will tell us the size of the continuum, but I cannot see how this property is discoverable.

5 The choice of axioms; discovering the truth

I have indicated that the formalist must choose the axioms, must decide which structures to study. The realist must seek a way of determining what are the ‘true’ statements about his real world. I will discuss how, in practice, the formalist makes his choice; it may well be that this method is philosophically naïve. My claim is that, mathematically, the process is very successful; I leave it to philosophers to decide how justifiable it is.

Ultimately the only binding constraint on the formalist’s choice of axioms is that they should be consistent, or at least that they should not be known to be inconsistent.⁹ This gives us a great deal of freedom. Nevertheless I am arguing that there are rational reasons, arising from the subject itself, that justify the consensus among mathematicians for the choice. The purpose of my examples was to exhibit some choices that have been made; I now seek to explain how this happened.

It appears to me that the realist has a far bigger challenge to justify how it is knowable that certain statements—however obvious, however useful—are ‘true’. Actually, the previous sentence is an euphemism. I cannot at all see how the significant mathematical statements that are the basis of our modern science—I am referring to statements about infinite sets—can be established as ‘true’; this is a problem for the realist, and I trust that I shall receive enlightenment on this point during this week. At present I shall take as my guide the procedure whereby Penelope Maddy (1993) seeks to determine whether or not the Axiom of Constructability is true; see also (Maddy 1998c).

In fact, it clearly emerges that the evidence that Maddy adduces in favour of the truth of a statement is very similar to that which I believe the formalist would adduce in favour of the choice of a particular scheme of axioms. Thus, in practice, the mathematical structures that are studied by formalists and realists (and naturalists—see (Maddy 1998c)) cannot be systematically distinguished from each other; the range of opinion within each sect on, say, the status of CH seems to be the same as that between the sects. This is the main reason why working mathematicians are indifferent to the

philosophical dispute between formalists and realists: whichever way the debate moves, mathematicians will basically still study the same structures. The objects of study, the style of work, the questions that are considered important will evolve with time (and there may be a lack of unanimity among practitioners), but this evolution will be driven not by philosophical debate, but rather by reasons that arise within the subject and by the pressure to find a mathematical framework in which to express the ideas that arise in physics and other sciences.

What then are the criteria that the formalist adopts in deciding on the axioms and definitions to be studied?

(I) *The first criterion is that axioms should be simple and clear, and should isolate the essential aspects of many diverse, known examples; the choice will have been successful if they are fecund in suggesting other, new examples, and in encompassing examples which arise in other contexts.*

The examples from which the axioms are abstracted will have arisen already in mathematics; they may be rather close to our physical perception (however unreliable) of the universe in which we exist, or they may have been abstracted from this perception, perhaps through several layers, so that the physical intuition lies far away.

For example, consider the definition of a ‘group’. Without going into a history of the long evolution of this ubiquitous concept, let me just point to the theory of permutations, Cauchy’s notion of the composition of substitutions of the early XIXth century, Galois’ study of the roots of equations of 1830,¹⁰ Hamilton’s quaternions of 1843, matrices, congruence classes in number theory, geometric transformations, etc., etc. Yet even Kronecker and Cayley, great algebraists of the XIXth century, did not work with the general notion of group: it seems that this emerged only around 1890. It is surely now universally recognized that the abstract concept of group is astonishingly successful, with a multitude of applications in science and elsewhere:¹¹ group theory is a pervasive language, now conquering new areas of physics, for example with the notion of a quantum group.¹²

The notion of Hilbert space arose from a desire to generalize that of finite-dimensional Euclidean space, namely, the spaces \mathbb{R}^n with the inner product

$$\langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = \sum_{j=1}^n x_j y_j.$$

This desire was fuelled by the need for a language to express important

physical concepts. The actual axioms arose in particular from contemplation of the two examples, H_1 and H_2 , that I mentioned above. They do seem to capture in a simple, clear way both our geometric and analytic concepts of ‘space’, allowing us the concepts of orthogonality and angle, but taking us beyond finitely many dimensions. In the hands of von Neumann, the theory of bounded linear operators on Hilbert spaces became, in the 1930s, the language for the new science of quantum physics.¹³ Thus Hilbert spaces have a rather direct application in physics. It is not asserted that an abstract Hilbert space ‘is’ a physical space arising in quantum theory, but that the language of Hilbert-space theory is a fruitful way of modelling the physical theory. The philosophical problem is why this theory is so unreasonably successful in this, allowing physicists to make predictions that are confirmed experimentally to an astonishing accuracy.

The rather complicated notion of a Banach algebra arises, not directly from physical concepts, but from the realization¹⁴ that the concept captures the essential features of many mathematical structures that have already been deemed to be significant; these include algebras of continuous functions such as $C(\Omega)$ and its subalgebras, convolution algebras arising in harmonic analysis and the theory of Fourier transforms, and algebras of bounded linear operators on a Banach space; in particular, the subclass of C^* -algebras includes the algebra of bounded linear operators on a Hilbert space already mentioned.

The above examples lie within mathematics, and philosophers may not be well-acquainted with, say, the theory of Banach algebras. But they are well-acquainted with the real line \mathbb{R} . I make the bold claim that the notion of ‘complete ordered field’ is very simple and clear and does indeed capture, at least from one perspective, our essential conception of what \mathbb{R} is. Further our formulation does suggest further examples: by dropping the requirement that the field be Dedekind complete, we encompass a plethora of examples, including the much-studied ultrapowers.¹⁵ I wonder if the realist would agree that it captures what is ‘true’ about the real line?

The final claim is that it seems that the axioms ZF or ZFC of set theory do capture our present intuition about sets; the axioms are so simple and clear that most mathematicians do not specifically mention that they are working in ZFC, and may not even realize it.

(II) *The second criterion that I apply is that of the depth of the development that takes place within the subject specified by the axioms.*

Consider the notion of a group. For a group G , a subgroup H is *normal* if $\{r \cdot s : s \in H\} = \{s \cdot r : s \in H\}$ for each r in G ; this is a fundamental concept, for normal subgroups are just the kernels of group morphisms. The group G is *simple* if the only normal subgroups are $\{e_G\}$ and G itself. Any classification theory of groups will seek to build an arbitrary group in some way from simple groups. So it is an immediate question what the finite, simple groups are. This is an easy question to ask, but formidably difficult to resolve: after decades of effort, the solution, giving a full list,¹⁶ is a triumph of our era. That there are developments of such depth within group theory by itself justifies the formulation of the concept. I claim that knowledge of the classification result, and more particularly the accumulation of techniques and understanding that led to the proof, enriches our subject.

(III) *The third criterion that I apply is the frankly aesthetic one.*

Mathematicians are willing to make such judgements. For example, justifying their book on Banach algebras, Bonsall and Duncan (1973, p. vii) write:

The axioms of a complex Banach algebra were very happily chosen. They are simple enough to allow wide ranging fields of application ... At the same time they are tight enough to allow the development of a rich collection of results ... Many of the theorems are things of great beauty, simple in statement, surprising in content, and elegant in proof.

The words ‘beauty’, ‘simple’, ‘surprising’, and ‘elegant’ are doubtless not easy to justify philosophically, but there is a wide consensus among mathematicians on how to recognize these attributes, and on how important they are.

(IV) *One should not arbitrarily restrict the notions under consideration unless forced to do so by the desire to avoid contradiction.*

This criterion is taken directly from Maddy’s argument against the suggestion that the Axiom of Constructibility be true. For example, I quote from Moschovakis (1980, p. 610):

The key argument against accepting [the Axiom of Constructibility] ... is that [it] appears to restrict unduly the notion of *arbitrary* set of integers; there is no a priori reason why every subset of ω should be definable ...

The argument is carried forward by Maddy (1993) with a discussion of the historical extension of the notion of function; through the centuries, there has been a movement to a more inclusive concept of function, so that I regard a function from S to T to be a subset R of $S \times T$ such that, for each s in S , there is a unique t in T with (s, t) belonging to R . Of course, other views, expressed in lectures here, would restrict the notion of ‘function’ to that which can be constructed in some way.

It is clear that the criteria that I have noted are subjective; they involve questions of judgement and experience within mathematics; the tests that are suggested will have answers that evolve with time; they are by no means uncontroversial, a very different view of the fourth criterion being taken by constructivists, for example. Perhaps they are aesthetic criteria. My argument is that the realist who seeks to justify the claim that his theorems are ‘true’ has no fundamentally more secure criteria for truth at his disposal.

6 Arguments against realism

My first argument against realism is clear: I do not see how we can know whether statements about the platonic set-theoretic universe are true or not. Arguments adduced for the alleged truth of the negation of the Axiom of Constructibility are convincing enough to lead me to make the aesthetic choice of not accepting this axiom; but I find them well short of compelling me to know the axiom to be a false statement about real sets. Maybe I will have been enlightened by the end of this conference! The extreme case is to convince me why various (very) large cardinals do or do not exist.

It has been suggested (Gödel 1947) that we shall resolve the size of the continuum because in time our understanding of sets will evolve to such an extent that eventually an ‘obviously true’ axiom about sets that resolves CH will be enunciated; I am very sceptical of this claim. Even if we do discover a very persuasive axiom that, *inter alia*, resolves the Continuum Hypothesis, and a majority of mathematicians absorb this axiom into their work, this does not make the axiom ‘true’. The Axiom of Choice is not ‘true’ because, in this century, the vast majority of practitioners have adopted it into their work, unless ‘true’ is defined by the last clause.

It has also been suggested that the questions whose truth we cannot resolve lie a long way from fundamental statements, and so an inner area

can be delineated in which we can readily recognize truth. This would seem to be an unsatisfactory procedure, even if possible. But the area of uncertainty encroaches on the heartlands. For example, it is now known that questions on the existence of large cardinals have influence on apparently elementary questions about \mathbb{R} . Consider the following example. It is not difficult to prove that $f(B)$ is a Lebesgue measurable subset of \mathbb{R} for each Borel subset B of \mathbb{R} and each continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$. But now suppose that f and g are continuous functions on \mathbb{R} . It is a remarkable fact that it cannot be decided in ZFC whether or not $f(\mathbb{R} \setminus g(B))$ is necessarily Lebesgue measurable, but, with the additional hypothesis that there is a measurable cardinal, all these sets are indeed Lebesgue measurable.¹⁷ Here is another example. We have remarked that $(C(\mathbb{I}), |\cdot|_{\mathbb{I}})$ is a Banach algebra. It was a famous question of Kaplansky whether any other norm $\|\cdot\|$ such that $(C(\mathbb{I}), \|\cdot\|)$ is a normed algebra is necessarily equivalent to the uniform norm $|\cdot|_{\mathbb{I}}$. It was eventually proved that, with CH, there are non-equivalent norms;¹⁸ it was assumed that this introduction of CH was a removeable blemish of the proof, but in fact it was proved by Solovay and Woodin that this result cannot be proved in ZFC.¹⁹ It is also known that there are models of ZFC + DC, where DC is the axiom of dependent choice, in which all sets of real numbers are Lebesgue measurable and all linear maps between Banach spaces are continuous.²⁰ Here we are seeing ‘real’ and fundamental questions from the perspective of analysts which cannot be resolved without precision about the set-theoretic axioms to be used. These are not isolated examples; they permeate the subject. I can give feeble indications why I prefer one resolution of these questions to another, but I cannot see how to determine the ‘true’ solution.

My second argument, as a glib formalist against realism, is that realism is restrictive. If it were known that a particular statement, such as CH, were true about \mathbb{R} , then no one could justifiably work in models in which the statement was false.

Let us suppose that we are working in ZF, and consider the two most basic independent axioms. The rôle of the Axiom of Choice (AC) was very controversial in the early years of this century,²¹ but it is now generally accepted by working mathematicians because, with this axiom, one can establish many results which we ‘wish’ to be true. For example, among the many facts that hold in ZFC, but which cannot be proved in ZF, are the following: each linear space has a basis; in a unital algebra, each proper ideal is contained in a maximal ideal; each field is contained in an algebraically

closed field; each filter on \mathbb{N} can be extended to an ultrafilter; Tychonoff's theorem; the Hahn–Banach theorem. We would feel unduly restricted without these facts at our disposal. Nevertheless the mathematics of the few who explore the consequences of $ZF + \neg AC$ is surely valid.²²

The balance of opinion about CH is more evenly divided, and I would not care to guess what the consensus, if any, will be in the future. The formalist position is strictly that any two relatively consistent extensions of ZFC are equally valid; I wish to know the theorems that arise in both $ZFC + CH$ and $ZFC + \neg CH$. Both sets of axioms lead to exciting mathematics; let both theories flourish!

7 Summary

I have explained, writing as a specimen of a working mathematician, that I am not unrepresentative of those who, if forced to make a decision, would call themselves *formalists*; that formalism is explainable in a 'precise, elegant, and self-consistent manner' that appeals to mathematicians; that we live our formal lives with rationally-chosen and enormously successful, albeit subjective, systems of axioms to which we have a real commitment; that this formalistic method has informed the great mathematical advances of the XXth century, and has become the dominant mode of exposition. It is unlikely that philosophical attacks at the level that we have experienced so far will drive us from our fertile fields whilst we are garnering such a rich harvest.

Notes

1. The view is taken by some mathematicians that mathematical logic and set theory are not part of their subject; perhaps even that it is to be dismissed from the canon of 'serious mathematics' because of its alleged lack of substantial content and its association with philosophy. It is very surprising to me that such views can be expressed, and I reject them; they must be based on ignorance. As a non-set-theorist, it is clear to me that the theorems proved within set theory in recent years are among the deepest, most technically sophisticated, and most significant within any area of mathematics.

2. For example, suppose that Ω is the set of all sets. Then every subset of Ω is a member of Ω , and so the power set $\mathcal{P}(\Omega)$ is a subset of Ω , a contradiction of a well-known theorem of Cantor.

3. Dummett (1994) distinguishes between *strict formalists*, who use first-order logic, and *semi-formalists*, who permit second-order logic.

4. An example of a metatheorem is the statement that the consistency of both of the theories $\text{ZFC} + \text{CH}$ and $\text{ZFC} + \neg\text{CH}$ follows from the consistency of ZFC ; see the Introduction for a discussion of the independence of CH .

5. Essentially our thoughts are derived from those of Hilbert and explained in §6 of the Introduction; but note the ‘internal tension’ in Hilbert’s view described in §4 of the Introduction.

6. Even today, the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

for the roots of the quadratic $ax^2 + bx + c$ is known to undergraduates; similar formulae for the roots of cubics and quartics were known from early modern times; however there can be no such general formula for the roots of quintic polynomials. ‘Galois theory’ makes this statement precise, and explains exactly why this is the case. For a popular account, see (Dieudonné 1992).

7. Formal definitions that generalize this idea arise in the branch of mathematics called *category theory*; see (Mac Lane 1971), for example.

8. Let A be an algebra which is semisimple, and suppose that A is a Banach algebra with respect to two norms $\|\cdot\|_1$ and $\|\cdot\|_2$. Then the identity map from $(A, \|\cdot\|_1)$ onto $(A, \|\cdot\|_2)$ is automatically continuous. This is Johnson’s uniqueness-of-norm theorem; see (Bonsall and Duncan 1973, 25.9).

9. Mathematicians are confident that, if an inconsistency were to emerge in, say, the axioms of ZFC , then a modest modification of the axioms would lead to a similar system without the inconsistency; this confidence can only be based on intuition and experience with the subject acquired by the community of mathematicians over two and a half millennia.

10. Just before he was killed in a duel at the age of 21, Galois laid out a general theory, based on the notion of a group, to the age-old problem of

when a general polynomial can be solved ‘by radicals’. For this theory, and a note on Galois’s life, see (Stewart 1989), for example.

The further history of Galois’s ideas is not without interest. In 1831, Galois submitted his memoir to the French academy; the referee, Poisson, declared it ‘incomprehensible’; it was not absorbed into general mathematical culture until the beginning of the XXth century; for many years, it has been a standard part of the undergraduate curriculum in England and throughout the world; and now, in the last two or three years, for reasons Effros (1998) would recognize, it seems to be disappearing from our curriculum because undergraduates find it ‘incomprehensible’.

11. The seminal rôle of group theory in the great physical theories of the XXth century, including relativity theory and quantum theory, is well-known; see, for example, the massive treatise (Cornwell 1984).

12. A ‘quantum group’ is not a type of group, but an *analogue* of a group, and so it is probably misnamed. The group algebra of a finite group and the enveloping algebra of a finite-dimensional Lie algebra have extra structure beyond their structure as an algebra (in the technical language they have a coidentity, comultiplication, and an antipode map); these extra structures make them into *Hopf algebras*. ‘Quantum groups’ are certain Hopf algebras. Whether all Hopf algebras ‘deserve’ to be called quantum groups is a matter for ongoing debate.

I explain the above for the following reason. Formally, a formalist studies the consequences of sets of axioms. But at this point in history the axioms to define a ‘quantum group’ may well not yet have evolved to a final form; there is genuine debate. The process of discussing which set of axioms most happily describes what we wish to call a ‘quantum group’ is a totally valid part of the life of a formalist; this is a period in which the criteria which I have suggested are being applied to delineate what will presumably within a few years become part of the canon – just as the concept of ‘group’ itself evolved in the last century.

13. See also the chapter of Jones (1998).

14. This was primarily by the great Russian mathematician, I. M. Gelfand, in the seminal paper (1941).

15. For an exposition of the theory of ‘super-real fields’, which are the natural generalization of the above concept of the real line \mathbb{R} as a complete ordered field, to ‘bigger’ real lines, see (Dales and Woodin 1996).

16. See (Solomon 1995) for a non-technical account. The proof of the classification theory was the work of very many mathematicians; in its present form, it could take 1000 pages for a full account, proving all the necessary intermediate results.

17. This example is taken from (Dales and Woodin 1996, p. viii).

18. See (Dales 1979) and (Esterle 1978).

19. The argument is the following. Consider the statement (NDH): for each compact space Ω , each norm $\|\cdot\|$ such that $(C(\Omega, \|\cdot\|))$ is a normed algebra is equivalent to the uniform norm $|\cdot|_\Omega$. We know from the result of Dales and of Esterle that NDH cannot be proved as a theorem in ZFC. We now start from the assumption that there is a model for ZFC, and, by a process of ‘forcing’, construct another model of ZFC such that NDH is a statement which holds in this model. (In this model, CH is necessarily false.) This establishes that the negation of NDH cannot be proved from ZFC, and hence that NDH is *independent* from ZFC. For an account of this proof, and a general exposition of forcing, see (Dales and Woodin 1987).

20. See (Solovay 1970).

21. For an interesting historical account, see (Moore 1982).

22. For an investigation into life without Choice, see (Jech 1973).

Bibliography

Bonsall, F. F. and Duncan, J. (1973). *Complete normed algebras*. Springer-Verlag, New York.

Bridges, D. S. (1998). Constructive truth in practice. *This volume*, 53–69.

Cornwell, J. F. (1984). *Group theory in physics*. Academic Press, New York.

Dales, H. G. (1979). A discontinuous homomorphism from $C(X)$. *American J. Math.*, **101**, 647–734.

Dales, H. G. and Woodin, W. H. (1987). *An introduction to independence for analysts*. London Math. Soc. Lecture Note Series, Vol. 115. Cambridge University Press.

Dales, H. G. and Woodin, W. H. (1996). *Super-real fields: totally ordered fields with additional structure*. London Mathematical Society Monographs, Vol. 14. Clarendon Press, Oxford.

- Davis, P. and Hersh, M. (1980). *The mathematical experience*. Birkhäuser, Boston.
- Dieudonné, J. (1992). *Mathematics—the music of reason*. Springer-Verlag, Berlin. Translated from *Pour l'honneur de l'esprit humain*. Hachette, Paris, 1987.
- Dummett, M. A. E. (1994). Reply to Oliveri. In *The philosophy of Michael Dummett* (ed. B. McGuinness and G. Oliveri), pp. 299–307. Kluwer Academic, Dordrecht.
- Effros, E. (1998). Mathematics as language. *This volume*, 131–45.
- Esterle, J. R. (1978). Injection de semi-groupes divisibles dans des algèbres de convolution et construction d'homomorphismes discontinus de $\mathcal{C}(K)$. *Proc. London Math. Soc.* (3), **36**, 59–85.
- Field, H. (1998). Which undecidable mathematical sentences have determinate truth values? *This volume*, 291–310.
- Gelfand, I. M. (1941). Nomierte Ringe. *Rec. Math. N.S. (Matem. Sbornik)*, **9**, 3–24.
- Gödel, K. (1947). What is Cantor's continuum problem? *American Mathematical Monthly*, **54**, 515–25. Expanded version in *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 470–85. Cambridge University Press, 1983. Reprinted in *Kurt Gödel: collected works*, Vol. 2 (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort), pp. 254–70. Oxford University Press, New York, 1990.
- Jech, T. (1973). *The axiom of choice*. North-Holland, Amsterdam.
- Jones, V. F. R. (1998). A credo of sorts. *This volume*, 203–14.
- Mac Lane, S. (1971). *Categories for the working mathematician*. Springer-Verlag, New York.
- Maddy, P. (1990). *Realism in mathematics*. Clarendon Press, Oxford.
- Maddy, P. (1993). Does V equal L ? *Journal of Symbolic Logic*, **58**, 15–41.
- Maddy, P. (1998c). How to be a naturalist about mathematics. *This volume*, 161–80.
- Martin, D. A. (1998). Mathematical evidence. *This volume*, 215–31.
- Moore, G. H. (1982). *Zermelo's Axiom of Choice: its origins, development, and influence*. Springer-Verlag, New York.