Truth in Mathematics

Edited by

H. G. Dales

Department of Pure Mathematics
University of Leeds

and

G. Oliveri

Wolfson College
University of Oxford

CLARENDON PRESS • OXFORD
1998

Oxford University Press, Great Clarendon Street, Oxford OX2 6DP
Oxford New York
Athens Auckland Bangkok Bogota Buenos Aires Calcutta
Cape Town Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul
Karachi Kuala Lumpur Madrid Melbourne Mexico City Mumbai
Nairobi Paris São Paolo Singapore Taipei Tokyo Toronto Warsaw
and associated companies in
Berlin Ibadan

Published in the United States
by Oxford University Press Inc., New York

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
(Data applied for)

ISBN 0 19 851476 X

Typeset using LaTeX
Printed in Great Britain by
Bookcraft (Bath) Ltd
Midsomer Norton, Avon

This volume is dedicated to the memory of Dr R. O. Gandy,

22 September 1919–20 November 1995.

Preface

The present volume is a record of lectures given at a conference *Truth in Mathematics*, which was held in Mussomeli, Sicily, Italy, from 13 to 20 September, 1995. All of the papers collected in this volume are based on lectures given at the conference; some are essentially verbatim records of what was spoken in Mussomeli, and some have been extended and somewhat changed.

Unfortunately, Professor Penelope Maddy was not able to attend the conference; we are grateful to her for sending a paper and to Professor Alexander George for reading this paper in Mussomeli. The paper of George and Velleman was also read by Professor George.

Stimulating lectures were given at the conference by Dr Daniel Isaacson and Professor Angus Macintyre, both of Oxford University. Unfortunately, papers based on these lectures are not included in the present volume.

The papers have been grouped into certain 'parts' of this volume; this is a necessarily rough classification. The notes to individual papers and their bibliography are collected at the end of each paper; further, there is a union of the separate bibliographies on pages 353–370.

Acknowledgements

Organizing an international philosophy conference is always a challenge both for the organizers and the participants. Cultural differences, discrepancies in expectations, and the actual situation on the ground are an inexhaustible source of wonderment and surprise.

Our meeting had a further twist to it with respect to similar events, a twist represented by the fact that the participants belonged to two rather different anthropological groups: some were philosophers, others mathematicians. But we must here acknowledge as one of the merits of all the participants that, in spite of all the potential sources of conflict present, they made the meeting at Mussomeli both useful and enjoyable. It even added to the fun that we could learn something of the cultural milieu of the 'other tribe'.

Special thanks must be given to the distinguished speakers, who established the scientific relevance of the event through their contributions, which have been collected in this volume. We are particularly grateful to Professor Michael Dummett of Oxford University for attending the conference and for acting as President. All the formal part of the conference (and much of the informal part) took place in English; we are most grateful to our Italian hosts for tolerating this and for providing charming and efficient translators between Italian and English. The reputation of the anglophones was somewhat redeemed by Professors Dummett and Jones, who made speeches of thanks in Italian; we are grateful to them for this.

Many people have helped us in making the Second International Philosophy Conference of Mussomeli a success. But, perhaps, the person who helped us most was Professor Roberto Prisco, who, with his enthusiasm and vision, inspired and accompanied us through all the many preliminary stages and who, through his personal involvement during the conference, contributed to the positive resolution of a considerable number of organizational problems.

Our thanks go to the Mayor, Professor Calogero Canalella, and to the Administration of the Town Council of Mussomeli for their generous funding of our initiative, which manifests in very clear terms their strong belief in the fact that culture at the highest level matters to the life of the community at large. Such an attitude is a breath of fresh air in a more and more stifling atmosphere created by today's pervasive, and all too often unchallenged, administrative wisdom, a wisdom which has sprung from purely ideological talk of 'efficiency gains', 'productivity' and 'profit'.

We wish to express our deepest gratitude to all those who dedicated their time and energy to the organization of the conference and, especially, to the people of Mussomeli who, once again, made us the object of their wonderful hospitality

showing how much they share in their Town Administration's aspirations and respect for culture.

We are also very grateful for the grants we received from the Philosophy Department and the Department of Pure Mathematics of the University of Leeds and from the Mind Association.

We are particularly grateful to Mrs Margaret Williams of the Department of Pure Mathematics at Leeds for dealing with the submitted papers, in their varying dialects of TEX, and for preparing the final copy for the publishers.

Finally, we must end on a sad note. It was a privilege and pleasure for us that Dr Robin Gandy of Oxford University attended the conference. We were of course all conscious of the great contribution that Robin made to the subject of mathematical logic in his distinguished career; Robin made many stimulating contributions to the discussions, and we believe that he greatly enjoyed the conference. We were deeply saddened to hear of Robin's death in Oxford on 20 November 1995, rather soon after the end of the conference.

Leeds H.G.D.
Oxford G.O.
February 1998

Contents

I Knowability, constructivity, and truth

II Formalism and naturalism

III Realism in mathematics

IV SETS, UNDECIDABILITY, AND THE NATURAL NUMBERS

1

Truth and the foundations of mathematics. An introduction

H. G. Dales and G. Oliveri

The concept of truth occupies a central position both in mathematics and in its philosophy. As we shall see in what follows, different conceptions of what it means to say of a mathematical statement that it is true give rise to different and mutually incompatible mathematical theories and to mutually exclusive positions on whether there are such things as mathematical knowledge and reality and on how these ought to be characterized.

The purpose of this introduction is to guide the reader to an appreciation of the importance of the theme of the book through: a preliminary clarification of the historical background relevant to the contemporary debate on the concept of truth in mathematics; a brief discussion of the mathematical and philosophical importance of such a concept; and a sketch of the applicability of the concept of truth in set theory.

Since some readers will be mathematicians lacking in philosophical background and others will be philosophers lacking in mathematical background, we have attempted, in writing this introduction, to define as many as possible of the important notions which are relevant to what we here say. We are confident of the fact that those who will read this introduction will be very much aware of the difficulty inherent in the task we have set ourselves to accomplish and that, therefore, they will treat in a benevolent way our efforts even when they appear to fall short of the target.

Explicit reference to chapters contained in this collection will be made whenever we touch upon a topic there treated in greater depth.

1 The pre-Tarskian debate: Kant

The origins of the debate on the nature of truth in mathematics lie in ancient antiquity. The Greeks—in fact, the school of Pythagoras, in the fifth century BC—discovered that, in our terms, there is no rational number x such that $x^2 = 2$. It follows that $\sqrt{2}$ is not a rational multiple of 1, and that there are segments (of length $\sqrt{2}$ and 1) which are incommensurable, a disturbing realization! Since the Pythagoreans did not go beyond the rational numbers, this discovery posed fundamental problems, challenging their central identification of number and geometry. In particular, they asked: What is the 'true' concept of number? Does $\sqrt{2}$ really exist? It will be seen that analogous questions are still relevant today.

However, since we cannot discuss the long history of the debate on 'truth' from the time of Greek civilization,[1] we move immediately to the XVIII$^{\text{th}}$ century of our era.

Contemporary speculations concerning the truth of mathematical statements originate from Kant's discussion of this problem, a discussion which results in the view that mathematical statements are synthetic *a priori*. But what does this mean?

For Kant judgements are expressed by statements which have a subject–predicate structure, that is, for example, by statements of the type 'John is bald'. An analytical judgement is, for him, a judgement expressed by a statement in which the predicate does not increase the amount of information contained in the concept of the subject, e.g., 1) 'The Sun is a body'—in this particular case the concept of body is implicit, *contained* in the definition of 'Sun'.

A synthetic judgement is, on the contrary, a judgement expressed by a statement in which the predicate increases the information contained in the concept of the subject, e.g. 2) '$\sqrt{2}$ is an irrational number' or 3) 'On the 21st of June 1995 Rome had more than one million inhabitants'. If someone were to give us a definition of Rome as the capital of Italy, etc., we would not be able to derive from this how many inhabitants Rome had on the 21st of June 1995; moreover, saying that $\sqrt{2}$ is the positive real number m such that $m^2 = 2$ does not give immediately any indication as to whether m is rational or irrational. Therefore, asserting that $\sqrt{2}$ is an irrational number and saying that Rome had more than one million inhabitants on the 21st of June 1995 are ways of *extending* the amount of information made available to us simply through an analysis of the definition of $\sqrt{2}$ or of Rome.[2]

Furthermore, for Kant, a judgement is true (false) *a priori* if it is true (false) independently of experience—the judgement expressed by 1) would be a good example of an *a priori* true judgement because we would be able to show that it is true simply through an inspection of the definition of 'Sun'; and that expressed by 2) would also be an example of *a priori* true judgement because we would be able to determine its truth by producing an abstract proof and not by experience.

Lastly, a judgement is *a posteriori* true (false) if only experience can justify the attribution of truth (falsity) to it: a clear example of *a posteriori* true judgement is represented by that expressed by 3). In fact, only a census taken in Rome on the 21st of June 1995, or a similar empirical verification procedure, is able to confirm or refute such a judgement.[4]

Kant's classification of judgements, and therefore statements, according to the analytic/synthetic and *a priori*/*a posteriori* distinctions allows us to construct a useful classification table, as in Table 1.

Table 1 Judgements classification table

	a priori	*a posteriori*
analytic	1)	
synthetic	2)	3)

In this table, the box individuated by the pair analytic/*a posteriori* is empty because, if a judgement is analytically true, then it is true in virtue of the fact that the predicate belongs to the concept of the subject. In other words, if we were to know the concept of the subject, that is, the list of predicates expressing properties of the subject, then, *independently of experience*, we would realize that the judgement is true, because we would find the predicate (occurring in the judgement) in our list. Therefore, such a judgement could not be *a posteriori*.

If, on the other hand, a judgement is genuinely *a posteriori*, that is for example, there is in principle no way of telling whether it is true or false independently of experience, then the judgement could not be analytic, because the list of predicates expressing the properties of the subject would neither contain the predicate occurring in the judgement nor a predicate which is inconsistent with it.

For Kant, mathematical judgements are synthetic *a priori*, because they are true independently of experience and, at the same time, extend the knowledge we can obtain simply by listing the properties of the subject.[4]

Although what we have so far said explains the meaning of the Kantian claim that mathematical judgements are synthetic *a priori*, we have said nothing about what Kant thinks true mathematical judgements are true of and about what, for him, is the relation between logic and mathematics. Both these questions need to be addressed, because they bear directly upon Kant's conception of the truth of a mathematical statement. The first, in fact, leads us to discuss his ideas on mathematical reality; the second helps us to clarify the nature of mathematical judgements by contrasting them with logical judgements.

In Kant's view, our reason is endowed with a spatio-temporal system of representation which produces a pre-reflective ordering of the perceptual input, that is, an ordering of the perceptual input that precedes the reflective activity of reason—this is the activity exercised by reason through the formation of judgements. What this means is that when we have a perception, this is not entirely determined in its properties and structure by the external object(s), but that such perceptory input is spatially and temporally ordered by reason. Moreover, such an ordering is pre-reflective because it is not the consequence of our attempt to interpret and/or understand what we are perceiving; it is simply given to us and determines, to an in principle unspecifiable extent, how things appear to us.

For Kant, the objects of study of mathematics are provided by this system of representation when certain *a priori* concepts are introduced; we are referring to concepts such as 'space is three-dimensional', 'given two points $\mathbf{a}$ and $\mathbf{b}$ in a plane α, there is one and only one straight line r in α on which $\mathbf{a}$ and $\mathbf{b}$ lie', 'adding two numbers m and n means ...', etc.[5] It is only because we have a spatial system of representation that, once we have introduced Euclidean geometry, we can discover, by studying instantiations of triangles and other geometrical figures, that 'the sum of the internal angles of a Euclidean triangle is 180°'. We could not have discovered this property of Euclidean triangles simply from an analysis of the definition of the *a priori* concept of triangle, because the analysis of an *a priori* concept always leads to the formulation of analytical judgements, and 'the sum of the internal angles of a Euclidean triangle is 180°' is a synthetic judgement.

Mathematical concepts must be *a priori* for Kant, and, therefore, cannot be *abstracted* from experience because of the necessity and universality of judgements obtained from their application; if mathematical concepts were empirical, the judgements in which they occurred would always have the contingency typical of judgements which can only be justified by induction.[6]

Moreover, for any concept and, in particular, for a mathematical concept to become meaningful, this has to refer to an object of our experience (an object given to us by our pre-reflective system of representation), hence the need to draw geometrical figures, write numbers, etc., that there is in mathematics.[7] One of the most important consequences of Kant's meaningfulness condition for concepts, when this is applied to mathematical concepts, is that it provides a strong justification for the belief that true mathematical statements *must* be true of something. [This statement is the 'principle C of Michael Dummett', which is discussed in the chapter of Martin-Löf (see pp. 105–114).]

Concerning the relation between logic and mathematics, it must be emphasized that, according to Kant, what he calls 'general logic', which is what in his system comes closest to what Frege and the following tradition meant by 'logic', is essentially concerned with the study of the laws of thought;[8] this is a study which is analytic in character because it:[9]

> ... resolves the whole formal procedure of the understanding and reason into its elements, and exhibits them as principles of all logical criticism of our knowledge.

Furthermore, despite the fact that this part of logic is seen by him as a propaedeutic to all the other sciences, it is not identifiable with them. Kant states that logic:[10]

> ... is justified in abstracting—indeed, it is under obligation to do so—from all objects of knowledge and their differences, leaving the understanding nothing to deal with save itself and its form. [...] for reason to enter on the sure path of science is, of course, much more difficult, since it has to deal not with itself alone but also with objects. Logic, therefore, as a propaedeutic, forms, as it were, only the vestibule of the sciences; and when we are concerned with specific modes of knowledge, while logic is indeed presupposed in any critical

> estimate of them, yet for the actual acquiring of them we have to look to the sciences properly and objectively so called.

Kant's views on the nature (synthetic *a priori*) and structure (subject/predicate) of mathematical judgements and on the relation between logic and some mathematical theories (arithmetic) have been strongly criticized by Frege. We will begin the next section with a brief account of Frege's opinions on the nature of mathematical judgements and on the relation between logic and mathematics.

2 The pre-Tarskian debate: Frege

Frege is regarded as the founder of modern mathematical logic; his views are expounded in his *Die Grundlagen der Arithmetik* (1884) and *Grundgesetze der Arithmetik* (1893, 1903).

There are several points of contact between Kant's and Frege's views of logic and mathematics. Both of them agree that logic is concerned with the study of the laws of thought;[11] mathematical concepts are not arrived at by abstraction; the study of mathematical judgements, conducted through the introduction of the *a priori*/*a posteriori* and analytic/synthetic distinctions, is very important to provide a correct characterization of mathematics; etc. Although these are certainly facts worthy of notice and of serious consideration, perhaps, the points on which Kant's and Frege's ideas about logic and mathematics are at odds with one another are even more relevant, at least to our general discussion, which, let us recall, revolves around the question 'How do we characterize the truth of a mathematical statement?'

For us, at the root of all the major differences between Frege's and Kant's representations of mathematics there are two irreconcilable conceptions of the relation between logic and mathematics. For Frege, contrary to Kant, there is no solution of continuity between logic and arithmetic; arithmetic *is* part of logic. However, this view of the relation between logic and arithmetic leaves him with a problem of difficult solution: if logic is, as Kant says, essentially analytic and arithmetic is part of logic, then arithmetic must be analytic as well. But, if this were the case, then arithmetic would lose the ampliative character possessed by a proper science.

Given the serious nature of the problem to be faced, Frege, to keep his views concerning the relation between logic and arithmetic, reformulates the definition of analyticity.[12]

For Frege, it is correct to say that arithmetical statements are analytically true (or false), because *their proofs* involve only laws of logic and definitions.[13] (From this follows, *a fortiori*, that logical judgements are analytic as well.) This move does the trick, for it allows Frege to say that arithmetical statements are analytical, preserving, at the same time, their ampliative nature of statements which, as Kant had already discovered, produce new information with respect to that contained in the concept of their subject.

According to Frege, the analytic/synthetic and *a priori*/*a posteriori* distinctions apply to *justifications* for making judgements, that is, to proofs, whereas, for Kant, they apply to judgements. Moreover, the proof of an arithmetical judgement is not constitutive of its truth, but is only a justification for asserting it:[14]

> It not uncommonly happens that we first discover the content of a proposition, and only later give the rigorous proof of it, on other and more difficult lines; and often this same proof also reveals more precisely the conditions restricting the validity of the original proposition. In general, therefore, the question of how we arrive at the content of a judgement should be kept distinct from the other question, Whence do we derive the justification for its assertion?

These last considerations show, in a very concise manner, how Frege attempts to hold on to his idea that arithmetic is just a part of logic without having to renounce the ampliative nature of arithmetical judgements. But they also, in particular those on the rôle of proof in arithmetic, coupled with Frege's belief in numbers as objects, which we have not discussed here,[15] make a case for attributing to Frege a realist account of the truth of arithmetical judgements.

However, as is well-known, despite the admirable array of arguments devised by Frege, the unsurpassable limitations of the logicist conception of the relation between logic and arithmetic were finally exposed when it became clear that logical notions alone are not sufficient to derive arithmetic.

Not only had Frege a compelling belief that arithmetic could be reduced to logical principles, but he also held that his base for such a reduction was secure. However, in a letter written to Frege shortly before Volume II of *Grundgesetze der Arithmetik* was published in 1903, Bertrand Russell showed that this was not the case, describing a logical contradiction that did arise: *Russell's paradox* considers the 'set of all sets that are not members of themselves'. Thus the naïve set theory which was part of Frege's logical basis is inconsistent in itself.

3 The pre-Tarskian debate: the rôle of Kronecker, Hilbert, and Brouwer

The relation between logic and mathematics, which has already provided us with a fruitful way of comparing Kant's and Frege's thoughts on the truth of mathematical statements, will also help us to introduce the seminal ideas of another author who has greatly contributed to such a debate: L. Kronecker.

Kronecker's approach to mathematics appears to be purely combinatorial. For him, mathematical activity is expressible as the ability to calculate; and such an ability can be acquired only once the natural numbers are given. According to Kronecker, as Marion puts it:[16]

> (1) everything must be construed from the natural numbers [...];
>
> (2) no completed infinities [are admissible within mathematics];
>
> (3) no proof of existence or definition [is given] without an algorithm.

In attempting to provide an explanation of the above principles (1)–(3), we must keep in mind that Kronecker's fundamental aim, which inspires the whole of his mathematical activity, is to proceed to the arithmetization of analysis and algebra. Seen in this light, principle (1) loses much of its ontological overtones, in the sense that it does not assert that the only mathematical entities are the natural numbers and that everything else has to be seen as a collection of illusions, but that we must endow each mathematical entity with an arithmetical significance through constructions starting from the natural numbers.

[For a discussion of two different conceptions of the natural numbers, see the chapter of George and Velleman on pp. 311–327.]

However, according to Kronecker, to achieve a rigorous arithmetization of analysis and algebra we must not presuppose the existence of complete totalities, because these are unnecessary—principle (2); in our reasonings, whenever we provide a definition or a proof of existence, we must also provide an algorithm which, in the case of a definition, enables us to show in a finite number of steps whether or not a given object belonging to the domain falls under it;[17] and, in the case of a proof of, say, $\exists x \varphi(x)$, enables us to produce, in a finite number of steps, an entity a belonging to the domain and to show that a has the property φ.[18]

[For a deep discussion of the notion of 'algorithm', leading to a significant proposal for 'founding the theory of algorithms'—this is part of the more general problem of 'founding computer science'—see the chapter of Moschovakis on pp. 71–104.]

Logical proofs of $\exists x \varphi(x)$, that is, arguments which appeal to the law of excluded middle, do not deserve to be called 'mathematical proofs'.[19] Weierstrass's arithmetization of analysis is not rigorous in Kronecker's eyes because, although Weierstrass succeeded in providing arithmetical significance to notions such as that of limit and continuous function, by means of his $\varepsilon - \delta$ notation and proofs, separating out these fundamental concepts of analysis from geometrical intuition, he disregarded the combinatorial character that *arithmetical* proofs must have to be logically compelling and made an indisciminate use of completed infinities.

However, for Kronecker, logical proofs are not sets of statements empty of any mathematical significance. They have an important rôle to play, not as justifications for mathematical assertions, but as heuristic devices which call for an *improvement* to be given in purely combinatorial terms.

These few remarks show that, for Kronecker, there exists a sharp divide between logic and mathematics and that mathematical truth, which still preserves a realist flavour, has to be given an arithmetical (numerical) significance. As we shall see in what follows, Kronecker's ideas greatly influenced the thought of Hilbert, Brouwer, and, later this century, Errett Bishop.

Another relevant contribution to the debate on the truth of mathematical judgements/statements—from here on we shall use these two terms interchangeably—is that given by David Hilbert.

In his work concerning the foundations of mathematics, Hilbert had to deal with two main problems: i) 'How do we justify the acceptance of Cantor's theory of transfinite numbers?' and ii) 'How do we deal with the set-theoretical paradoxes?'. (We shall describe Cantor's theory and the paradoxes in §6, below.)

Both these problems have an important characteristic in common which was well individuated by Hilbert: the concept of infinity. In fact, Cantor's theory of transfinite numbers might be described as an extension of arithmetic beyond the realm of the finite; and set-theoretical paradoxes such as Cantor's, Russell's, and Burali-Forti's are sensitive to contexts in which restrictions are placed on the size of totalities which we are allowed to call 'sets'.[20]

Perhaps, in the light of such a characteristic common to the two questions, we might be able to reduce i) and ii) to (a) 'Is there a way of treating the concept of infinity that might enable us to retain the results obtained by Cantor in transfinite theory of numbers, but avoiding, at the same time, the formation of the set-theoretical paradoxes?'.

Hilbert attempted to provide a positive answer to this question by developing what became known as *Hilbert's programme.*[21] But, before we proceed to illustrate Hilbert's views concerning how question (a) should be addressed, we need to set in place some important ideas.

For Hilbert, given a mathematical theory such as number theory, we can distinguish between *finitary* and *ideal* statements. Finitary statements are intuitively characterizable as those (and only those) whose truth can be assessed by an actual procedure of computation which takes place in a finite number of steps.[22] All the other statements are ideal. An example of finitary number-theoretical statement is provided by '$1 + 355 = 356$'. In fact, in this case there exists a well-specifiable computational procedure which, in a finite number of steps, is able to determine whether the statement is true or false.[23] Quantified number-theoretical statements which are not expressible as finite conjunctions or finite disjunctions of finitary statements are ideal. The reason for this is that an infinite conjunction (or disjunction) of number-theoretical statements is not something whose truth value can be determined by a computational procedure which terminates in a finite number of steps, but only by arguments which are essentially non-combinatorial—here the influence of Kronecker's ideas ought to be evident!

The distinction between finitary and ideal statements of a given mathematical theory is very important because all true finitary statements can be effectively shown to be so (true) and, therefore, the finitary fragment of the given mathematical theory must have a model and hence must be consistent.

According to Hilbert only finitary statements are meaningful, that is, true or false. Ideal statements, on the other hand, despite their lack of meaning, turn out to be *useful* to the development of mathematics, because they help us to preserve the laws of logic and mathematics in their simplest form.[24] The only requirements for ideal statements to be accepted as legitimate members of mathematical theories are their mathematical usefulness and the fact that they do not generate contradictions when adjoined to the finitary fragment of the given theory.[25]

Hilbert's programme consisted essentially in the attempt to show that the formalized version of a given mathematical theory is consistent by using a metatheory which appeals exclusively to finitary statements and reasoning. If the programme could be effectively carried out, it would indeed provide a positive and satisfactory answer to question (a). In fact, if Cantor's theory of transfinite numbers could be conceived as an extension of arithmetic and we were able to prove the consistency of a formalized version of arithmetic by means of a finitary meta-theory of arithmetic, then this consistency proof would be compelling—in force of its combinatorial character—and would show that we have nothing to fear from being admitted to Cantor's paradise!

As is well-known, even if some authors have recently disputed this,[26] the original version of Hilbert's programme came to grief as a consequence of Gödel's second (or incompleteness) theorem; see §6, below.[27] But what is of interest to us here is Hilbert's belief that only finitary (verifiable) mathematical statements are meaningful (true or false), and that mathematics has a subject matter represented by symbols and their immmediately clear and representable structure.[28]

However, Hilbert's view is strained by an internal tension generated by two beliefs which seem to be pulling in opposite directions. On the one hand, we have the opinion that only finitary statements are meaningful (true or false); which seems to imply that the predicate 'true' ought to be understood as a synonym of the predicate 'combinatorially provable'. From this it would, in particular, follow that the use of the predicate 'true' would not suggest any reference to a subject matter. On the other hand, Hilbert's appeal to an extra-logical reality, even though this is a reality populated by structured symbols, and his holding on to the law of excluded middle, that is, to logical proofs in Kronecker's sense, seem to hint at the possibility that it is not after all correct to take 'X is true' to mean 'X is combinatorially provable'.

As we shall see in what follows, such a tension is resolved in the thought of the intuitionists by keeping the belief that only finitary mathematical statements are meaningful and dropping, together with the law of excluded middle, any hint at a subject matter of mathematical theories.

[A position that is more radical than that of Hilbert is developed by Effros in his chapter on pp. 131–145. Effros sees mathematics as a language, discussing this language as a vehicle for expressing our ideas of the physical universe. A defense of a purely formalist view—namely, that mathematics is a mental game with strict rules—may be found in the chapter of Manin, pp. 147–159.]

The last of the thinkers who contributed to the pre-Tarskian debate on the truth of mathematical statements that we are going to consider here is L. E. J. Brouwer, the prophet and high priest of Intuitionism in the philosophy of mathematics.

For Brouwer saying 'A is true' means 'There is a constructive proof of A',[29] but what do we have to understand by 'constructive proof' or 'construction', in Brouwer's sense? Well, the answer to this question cannot be completely straightforward, because, although Heyting produced an interpretation of the logical constants, which is currently accepted as a fair embodyment of the 'constructive' ideas about intuitionistic first-order logic (see note 45), the Brouwerian notion of mathematical construction, as involving vastly more than logic, has been taken as primitive by the faithful followers of Brouwer. The reason for this is that, since constructions are creative acts of the subject, we can distinguish between them, as van Stigt says, 'genetically', but we cannot provide a complete classification of constructions. This situation is determined not so much by the fact that *there are too many* of them (constructions) or by the inadequacy of ways of formalizing constructions, but by the *openness* of the notion of construction, which is made to depend on the vague notion of 'creative activity' of the subject.[30]

However, to try to make the situation concerning the notion of construction more precise, we can say that, at the level of what we called 'finitary reasoning', intuitionists and followers of Hilbert's programme are in complete agreement, that is, finitary reasoning *is* constructive—this, in particular, shows that Brouwer was also influenced by Kronecker's ideas.[31] The differences between Brouwer's approach and that of the Hilbertians begin to arise at the stage in which ideal statements (in Hilbert's sense) are called into question. These last are deemed to have *no* mathematical value by Brouwer who, by adopting a more severe view on these matters than Kronecker, courageously attempted to *extend* the finitary standpoint to the treatment of any meaningful mathematical statement.

Here the extension of the finitary standpoint in mathematics must be interpreted in the sense that: i) we can consider as legitimate those operations which can be *in principle*, but not necessarily actually, carried out (they do not halt after a finite number of steps); ii) we are allowed operations which are not determinate in advance by a *law*; iii) sentences which do not lend themselves to be treated by such finitary methods are *not* part of mathematics at all.

The rationale for the extension of the finitary standpoint expressed by i) and ii) is represented by the need to deal with infinite structures and, in particular, with the continuum. Unfortunately, there is not enough space here to justify this assertion; the interested reader may consult on this standard texts, such as Dummett's *Elements of Intuitionism*.[32] Concerning the motivation for the adoption of principle iii), we need to consider that:[33]

> The point is not just that the intuitionist prefers constructive proofs to a greater degree than other mathematicians. A classical mathematician may spend a considerable amount of time looking for a constructive proof of a result for which he already has a non-constructive one. The intuitionist is not in this position; he must have a constructive proof because the intuitionistic interpretation of the conclusion is always such that no non-constructive proof could count as a proof of it.

The consequences of this attitude are momentous. The law of excluded middle is rejected, together with much of the mathematical practice and results achieved through the classical approach.[34]

One of the most important consequences of the fact that, in Brouwer's system, the predicate 'true' collapses on to the predicate 'constructively provable' is that when we assert 'X is true', we do not need to refer to the existence of a reality that X is true of.

Brouwer's intuitionism and other constructive approaches have had considerable philosophical impact in this century, but only a small number of mathematicians have espoused the cause of 'doing intuitionistic or constructive mathematics'. The most important of these mathematicians was Errett Bishop. [For our account of modern constructive mathematics, see the chapter of Bridges, pp. 53–69.]

The situation concerning what we must understand when we say that a mathematical statement is true remained rather fluid until Tarski's contributions to the debate, contributions which we shall examine in the next section.

4 Tarski

Tarski's work on truth[35] represents a watershed in the understanding of what it means to say that a statement is true in formalized languages.[36] As is well-known, Tarski held a semantic conception of truth—based on the concept of satisfiability—aiming at capturing the Aristotelian intuition that:[37]

> The truth of a sentence consists in its agreement with (or correspondence to) reality.

For Tarski a definition of truth that captures such an intuition must *imply* all the equivalences of the kind:[38]

1) 'The cat is on the mat' is true if and only if the cat is on the mat.

We shall not attempt to rehearse here a very well-known theory; the interested reader should have no difficulty in accessing the relevant literature. What is of interest to us is to individuate the factor present in Tarski's ideas that caused the traditional debate on the truth of mathematical statements to change.

The decisive novelty in Tarski's ideas is not contained in his account of what it means to say of a mathematical statement that it is true; as we have already seen, such a view is but a rediscovery of one of Aristotle's intuitions on these matters. The new and crucial move made by Tarski is rather part of his successful attempt to dispose of the semantic paradox of the liar, a paradox expressed by sentences such as:

2) This sentence is false.

Such sentences turn out to be true if and only if they are false, generating contradictory statements.

According to Tarski's analysis, two conditions must be satisfied to generate the liar paradox: i) we must be dealing with sentences belonging to a *semantically closed* language,[39] and ii) we must use classical logic. Tarski's strategy in dealing with the liar paradox consists in keeping classical logic and dispensing with semantically closed languages. The elimination of semantically closed languages is obtained by means of the introduction of the well-known distinction between object language and meta-language.[40]

One of the consequences of Tarski's monster-barring solution is the relativization of the definition of truth to a particular language L. It is such a relativization of the notion of truth which determines an important shift with respect to the past. Until Tarski, thinkers in the philosophy of mathematics and logic worked with an unrelativized, *absolute* notion of truth of a statement.

[A defense of the Tarskian conception of mathematical truth, based on an account of mathematical experience, may be found in the chapter of Oliveri on pp. 253–269.]

From the time of the appearance of Tarski's theory of truth onwards, the discussion concerning how we should understand the claim that certain mathematical statements are true became polarized between two main parties: those who accepted Tarski's account—we shall call them 'classical mathematicians' and will refer to classical mathematics by the term CLASS—and the constructivists, that is, the followers of Brouwer (INT) and of other mathematicians such as Markov (RUSS), Bishop (BISH), and Yessenin-Volpin, who produced systems of constructive mathematics which differ from Brouwer's.[41]

A simple example will, perhaps, be sufficient to highlight the distinction between the classical and constructive interpretation of a statement being true. Let us consider the following *Tarskian* definition of truth for a first-order language L.[42]

Definition 4.1 *A well-formed formula* $\mathcal{A}$ *is* true for the interpretation *or* model $\mathfrak{M}$ *(written* $\models_{\mathfrak{M}}$*) if every sequence in* Σ *satisfies* $\mathcal{A}$*. The formula* $\mathcal{A}$ *is said to be* false for $\mathfrak{M}$ *if no sequence in* Σ *satisfies* $\mathcal{A}$*.*[43]

Now for a classical mathematician to prove that a well-formed formula $\mathcal{A}$ of a first-order language L is true for $\mathfrak{M}$, it is sufficient to show that $\neg\neg\mathcal{A}$ is true for $\mathfrak{M}$. The reason for this is as follows. We can prove from our definition that:[44]

(*a*) $\mathcal{A}$ is false for an interpretation $\mathfrak{M}$ if and only if $\neg\mathcal{A}$ is true for $\mathfrak{M}$;

(*b*) $\mathcal{A}$ is true for an interpretation $\mathfrak{M}$ if and only if $\neg\mathcal{A}$ is false for $\mathfrak{M}$.

Thus, if we can prove that $\neg\neg\mathcal{A}$ is true for $\mathfrak{M}$, then this will imply, by (*a*), that $\neg\mathcal{A}$ is false for $\mathfrak{M}$; and this last result will imply, by (*b*), that $\mathcal{A}$ is true for $\mathfrak{M}$.

However, for the constructivist mathematician the argument we have given above is not at all sufficient to show that $\mathcal{A}$ is true because it has not provided us with a procedure to transform a *proof* of $\neg\neg\mathcal{A}$ into a *proof* of $\mathcal{A}$; in fact, the constructivist mathematician would be prepared to hold $\mathcal{A}$ to be true just in case he can produce a constructive proof of $\mathcal{A}$.

A good insight into the approach to truth in terms of constructive provability may be given by studying the constructivist's interpretation of the logical constants. This is an interpretation which greatly differs from that of CLASS. Following Troelstra and van Dalen, we shall call this interpretation the BHK-interpretation (or the Brouwer, Heyting, Kolmogorov interpretation).[45]

However, the difference between CLASS and the various types of constructive approaches to mathematics does not end here. In fact, as a consequence of the BHK-interpretation of the logical constants, INT, RUSS, and BISH do not accept the law of excluded middle $P \vee \neg P$[46] and, moreover, analysis as developed in

INT and RUSS is inconsistent with analysis as developed in CLASS.

Some authors, including Bridges in his chapter in this volume, dispute this last point. They assert that certain results which are provable in INT analysis and are classically false are only apparently so. This impression vanishes, they say, when we correctly state them. Bridges gives the example of the theorem of INT analysis stating that:

> (*) Every function from $[0,1]$ to $\mathbb{R}$ is uniformly continuous;

the apparent absurdity of (*) (and its inconsistency with CLASS analysis) disappears when we re-state it more carefully as:

> (**) Every intuitionistically defined function from the intuitionistic interval $[0,1]$ to the intuitionistic real line is, intuitionistically, uniformly continuous.

We do not intend to contribute here to the debate on this issue. Bridges (and others) may well be right, but, if this were the case, then INT and RUSS, far from being competing interpretations of current mathematical practice, interpretations which pose a strong challenge to CLASS (and to each other), would turn into mutually exclusive and incommensurable activities for which justifications ought to be provided to call them 'mathematical'.

However, regardless of what the situation is with respect to this issue, the differences existing between CLASS, on the one side, and INT, RUSS, BISH, etc., on the other, in the interpretation of what it means to say that a mathematical statement is true have profound metaphysical and epistemological consequences which we shall briefly discuss in the last section. In the next section we shall set out the background for our discussion of the notion of truth in set theory.

5 Truth in set theory: preliminaries

We now expand a little on the notion of model, introduced above. [A more detailed version of these ideas, leading to a formulation of Gödel's completeness theorem, is given in Woodin's chapter in this volume, pp. 329–351.] The concept of 'truth', in the sense to be described, informs most of twentieth-century set theory, and hence, implicitly most of the mathematics of our century.

We must start with a 'language' of set theory; we are thinking of 'first-order language, with equality', but other, more general, languages are possible. The *alphabet* of the language is an infinite set of positive integers, some of which are denoted by $\vee$, $\neg$, (,), $\exists$, and $=$; some finite sequences in the alphabet are the *formulae* of the language. The *syntax* specifies which sequences are formulae: it is a specific list of rules which state how formulae are formed (for example: 'if φ is a formula and x is a variable symbol, then '$\exists x(\varphi)$' is a formula'.) The language is the smallest set of finite sequences which is closed under applications of the rules. A *sentence* is a formula which has no free variables; intuitively, a sentence is an assertion that is either true or false. A *theory* (in the language) is a (possibly infinite) set of sentences.

How does a theory 'prove' a formula? The language contains certain *logical axioms* (for example, '$\varphi \to (\psi \to (\varphi \wedge \psi))$' is regarded as trivially true). The classical mathematician takes '$\varphi \vee (\neg\varphi)$' as a logical axiom; as explained above, it is this controversial act that fundamentally separates the classical mathematician from the constructivist. A *rule of inference* is a procedure for deriving a new formula from an existing collection. *Modus ponens* is the rule:

$$\text{from } \{\varphi,\, \varphi \to \psi\}, \text{ derive } \psi\,.$$

A theory T *proves* a formula φ in the language if there is a finite sequence $\langle \varphi_1, \dots, \varphi_n \rangle$ of formulae such that $\varphi_n = \varphi$ and each φ_i (for $i = 1, \dots, n$) is either an element of T or a logical axiom or is derived from two formulae of $\langle \varphi_1, \dots, \varphi_{i-1} \rangle$ by modus ponens. In this case we write $T \vdash \varphi$. Note that this definition does not imply that there is an algorithm that determines whether or not $T \vdash \varphi$.

Let T be a theory. Then T is *inconsistent* if there is a sentence φ such that $T \vdash \varphi \wedge (\neg\varphi)$; otherwise, T is *consistent*, and we write

$$\operatorname{Con} T\,.$$

A sentence φ is *relatively consistent* with T if

$$\operatorname{Con} T \quad \text{implies} \quad \operatorname{Con}(T + \varphi)$$

(that is, either T is inconsistent or $T + \varphi$ is consistent).

A *model* $\mathfrak{M}$ (of our language) is just a pair (M, E), where M is a non-empty set and E is a subset of $M \times M$.

Let $\mathfrak{M}$ be a model, let $\varphi = \varphi(x_1, \dots, x_n)$ be a formula of our language, and let $a_1, \dots, a_n$ be elements of M. There is a natural interpretation of φ as a statement (relative to $\mathfrak{M}$) about the elements $a_1, \dots, a_n$; the *truth* of φ at $(a_1, \dots, a_n)$ is defined by various natural rules. In particular we have the notion of a sentence φ being true in $\mathfrak{M}$; we write $\mathfrak{M} \models \varphi$ in this case. The *theory* of $\mathfrak{M}$, written $\mathrm{Th}(\mathfrak{M})$, is the collection of sentences φ such that $\mathfrak{M} \models \varphi$; $\mathfrak{M}$ is a *model* of a theory T if $T \subset \mathrm{Th}(\mathfrak{M})$, and we then write $\mathfrak{M} \models T$.

It is easy to see that, if a theory T has a model, then T is consistent. The foundation of the subject is the converse: it is *Gödel's completeness theorem.*[47]

Theorem 5.1 *Let T be a theory in our language. Then T is consistent if and only if T has a model.* □

It follows from Gödel's completeness theorem that a sentence is logically derivable if and only if it is true in every model. More generally, for a theory T, $T \vdash \varphi$ if and only if each model of T is also a model of φ. This is the importance of the theorem: it identifies the notions of proof and of truth.

[The chapter of Lolli (pp. 117–129) is an account of the notion of truth that stresses the fundamental rôle of Gödel's completeness theorem.]

[A related, but distinct, notion of proof is that of 'verificationism', which derives from the ideas of Gentzen, and is explained in the chapter of Prawitz (pp. 41–51); Prawitz concludes by claiming 'we should identify the truth of a sentence not with it being proved but with it being provable'. Gentzen's work is also discussed in the chapter of Moschovakis, pp. 71–104.]

Naturally, we would like our theories to be consistent. Another apparently desirable feature of a theory is that it be complete. Let T be a theory in a language. Then $\mathrm{Thm}(T)$ denotes the set of sentences φ such that $T \vdash \varphi$. A theory T is *complete* if, for each sentence φ, either $\varphi \in \mathrm{Thm}(T)$ or $\neg\varphi \in \mathrm{Thm}(T)$. A sentence φ is *independent* of T if $\varphi \notin \mathrm{Thm}(T)$ and $\neg\varphi \notin \mathrm{Thm}(T)$. Thus φ is independent of T if T proves neither φ nor $\neg\varphi$; if T is consistent, then φ is independent of T if and only if both φ and $\neg\varphi$ are relatively consistent with T. Certainly there are modest theories that are consistent and complete: for example, the theory for a dense total order without end-points has these two properties. But we would like more significant complete and consistent theories. The search for such theories is described in the following section.

6 Truth in set theory: axiomatics and independence

The background to Hilbert's questionings lay in the tumultuous development of new ideas in set theory in the latter part of the nineteenth century. These new ideas were primarily due to Cantor.[48] After Cantor's work, which we discuss below, smouldering controversies broke into the open, and the question of the proper foundation of mathematics became more insistent.

A crucial rôle in the debate about mathematical truth is played by the continuum hypothesis (CH), a matter first recognized and formulated by Cantor. Let us explain this statement.

Let S and T be two sets. Then S and T are *equipotent* if there is a bijection from S onto T; we also say that S and T have the same *cardinality.* Cantor studied the cardinality of sets (he used the term 'power of a set'), and he quickly singled out two cases of infinite set: *denumerable* sets, which are equipotent to $\mathbb{N}$, and *non-denumerable* sets, which are equipotent to the set $\mathbb{R}$ of all real numbers. Already in 1874, he had shown that $\mathbb{R}$ is not equipotent to $\mathbb{N}$ and had also proved that, in our terminology, $|S| < |\mathcal{P}(S)|$ for each set S, where $|S|$ is the cardinality of S and $\mathcal{P}(S)$ is the power set of S.

However, Cantor could not decide whether or not there is a set T such that $|\mathbb{N}| < |T| < |\mathbb{R}|$. In the standard modern formulation (whose notation derives from that of Cantor), we have: $|\mathbb{N}| = \aleph_0$; the next cardinal larger than $\aleph_0$ is $\aleph_1$; and $|\mathbb{R}| = 2^{\aleph_0}$. We know that $\aleph_1 \leq 2^{\aleph_0}$, but we cannot determine whether or not $\aleph_1 = 2^{\aleph_0}$. *The continuum hypothesis* is the statement:

$$\aleph_1 = 2^{\aleph_0}\,.$$

(There is also a generalized continuum hypothesis, namely $\aleph_{\alpha+1} = 2^{\aleph_\alpha}$ for each α. This statement is explained in numerous books.) It was a basic dogma of Cantor that CH is true.

Cantor was also troubled by whether or not every set could be well-ordered, and thereby associated with one of his aleph numbers as its cardinal; he did not know how to deal with the 'set of all sets'. Russell's paradox exhibited the intrinsic untenability of this latter concept. A magisterial attempt to preserve the essence of Frege's approach, avoiding the danger of paradoxes, was the *Principia Mathematica* of Russell and Whitehead (1910–13). This work develops an elaborate theory of 'types'; antinomies are avoided by enforcing a 'vicious-circle principle' that no set may be defined by reference to a totality that contains the set to be defined. [A modern view somewhat related to this programme, is given by Tait in his chapter, pp. 273–290.]

However, to most mathematicians the *Principia* must have seemed a desolate undertaking. Hilbert did not follow this approach; he certainly did not wish to be driven out of the 'paradise' that was Cantorian set theory. His aim was to make the foundations of set theory secure, and he led the new advance into axiomatics.

Hilbert had already had a successful approach to the axiomatic foundation of geometry in his *Grundlagen der Geometrie* of 1899. The next part of his programme was to set out the primitive concepts of arithmetic and to determine relations among these concepts by introducing appropriate axioms. The key feature of the approach was to establish that the resulting system was consistent and complete: that no contradiction could result from any combination of axioms, and that the system of axioms was sufficient to prove all theorems concerning the real numbers; in particular CH would be confirmed.

Hilbert's axiomatization of the real number system was followed by Zermelo's attempt to do the same for set theory itself. This attempt was the great focus of the debate on what is true in mathematics in the first part of this century. We have for example, the great mathematician Poincaré arguing that most of the ideas of Cantorian set theory should be banished from mathematics; that knowledge of mathematical entities originated in intuition and that they were thus *synthetic a priori* judgements in the Kantian sense; that Russell and logicism should be opposed because then mathematics would be nothing more than a system of tautologies. Versions of these ideas can be found in this volume.

Nevertheless, the wave of axiomatization surged forward, and in due course a fairly canonical collection of axioms of set theory emerged. This is the ZF = Zermelo–Fraenkel system, eventually enlarged by the addition of the Axiom of Choice AC to become the system ZFC. (There are other systems of axioms, such as GB = Gödel–Bernays set theory.) The inclusion of AC in this list was very controversial in its day, but the dust has probably settled on this dispute now; almost all working mathematicians do freely use, implicitly or explicitly, this axiom in their work.[49] [A list of the axioms of ZFC, together with some discussion, is given in §2 of Woodin's chapter in this volume (pp. 329–351); of course there is a multitude of texts containing this list.]

[A view that is a descendant of that of Hilbert is espoused by Dales in his chapter of this volume, pp. 181–200.]

The first question to be asked in connection with this approach to the foundations of mathematics is whether or not ZF or ZFC are consistent, that is, whether or not they have a model; the second is whether or not they are complete. These are most important questions. It was widely hoped and expected that the answer to both questions would be 'yes'.

However, there is a less precise question about ZFC that is very relevant to the nature of truth. If the axioms of the theory ZFC are simply an attempt to capture basic intuitions about sets, a judgement on the success of the enterprise must be subjective. But is this really the case? In other words, are there 'true' statements about sets that the axioms try to capture, or do the axioms *define* what is 'true' about sets?

Consider the question of whether ZFC is complete. If this were so, mathematics as we know it would reduce to a 'finite search'. However, happily, this is not the case. The bombshell that shook the axiomatic approach to its foundations was the announcement of Gödel in 1931 of his *incompleteness theorems.*[50] The first incompleteness theorem states that, if T is a consistent theory containing arithmetic (a subset of ZF) and such that the set of axioms of T is 'recursive', then there is a sentence which is independent of T. Thus a consistent theory cannot be complete!

The sentences manufactured by the proof of the first incompleteness theorem to be independent of ZFC are neither exciting nor natural. One might hope that all 'natural' sentences are not independent; in this case, the failure of Hilbert's programme would not seem to be very important. However, let us consider the hypothesis CH; most people regard CH as a 'natural' statement about the real line $\mathbb{R}$, and some regard it as 'exciting'. But we can now prove that CH is independent of ZFC.

The method of proof that has been most successful in this area is *algebraic.*

Let φ be a sentence; we seek to show that φ is relatively consistent with ZFC. Our assumption is that ZFC is consistent, and so has a model, and we seek to build a model of ZFC + φ. Gödel followed this method in 1938, when he proved that Con ZFC implies Con (ZFC + CH); under the assumption that ZFC has a model $\mathfrak{M}$, there is a submodel $\mathfrak{N}$ of $\mathfrak{M}$ such that $\mathfrak{N} \models \text{ZFC} + \text{CH}$.

In essence, Gödel discovered an effective method of building submodels of a given model. The actual submodel that he used is called the *constructible universe.* It was not until 25 years later that a method of building *extensions* of a given model was found; the technique of *forcing*, which was discovered by Cohen in 1963, allows one to extend models by adding new elements in a controlled manner; the sentences true in the extended model are exactly those which are forced to be true.[51]

The method of forcing, as developed by Cohen, proves that Con ZFC implies Con (ZFC + ¬CH); under the assumption that ZFC has a model $\mathfrak{M}$, there is an extension $\mathfrak{N}$ of $\mathfrak{M}$ such that $\mathfrak{N} \models \text{ZFC} + \neg\text{CH}$. Thus CH is independent of ZFC. Similarly, it is shown that AC is independent of ZF.

Gödel's original approach to his first incompleteness theorem was not algebraic, but *number-theoretic.* Formulae are coded and acquire a certain *Gödel number.* Thus the question of the independence of certain sentences from a theory can be reduced to the question of whether or not the Gödel numbers of certain formulae are in the range of a certain polynomial in k variables on the set $\mathbb{N}^k$. Highly relevant here is Matijacevič's solution to Hilbert's tenth problem: it is not possible to find an algorithm for testing an arbitrary polynomial equation $p(X_1, \ldots, X_n) = 0$, where p has coefficients in $\mathbb{Z}$, for the possession of a solution in integers. (One can take n to be 9, and p could be written down quite explicitly if one had sufficient fortitude.) The point of these remarks is to stress the fact that statements about the independence of CH can be transformed into statements about the solutions of polynomials, and that all these latter statements involve only the integers, the most basic building blocks of our science.

Nowadays a vast number of natural statements are known to be independent of ZFC; indeed it is rather a surprise to find a statement that is not so independent. An extensively played game is to take a statement, to decide whether it or its negation can be proved in the theory ZFC, and then, failing that, to decide whether the statement or its negation (or both) are relatively consistent with ZFC.

Thus we come to a basic challenge to philosophies of truth in mathematics: what does it mean to say that CH, and other independent statements, are 'true', given the theorems described above? This question is explored in several chapters in this volume. For example, Field (pp. 291–310) explains that 'The usual platonist view is that ... there is still a serious question as to whether [CH] is true, and we can still find indirect evidence for its truth. The plenitudinous platonist view is that there is no such question ...'. Field seeks to show that at least the concept of 'finite' can be given a determinate truth value, and hence that '... every number-theoretic sentence gets the same truth-value in every allowable model ...'. Again, Martin (pp. 215–231) discusses what could count as evidence for a mathematical truth. The two interesting examples he discusses go some way to justifying various 'determinacy axioms' that go beyond ZFC. (These determinacy axioms have been shown, by some very deep, sophisticated, and technical recent work of Martin, Steel, and Woodin, to be essentially equivalent to various 'large cardinal axioms'; these results, which were very surprising and not contemplated until rather recently, again show that theorems proved within set theory can throw light on the philosophical debate about the nature of mathematical truth.) A starting point for the chapter of Maddy (pp. 161–180) is also the fundamental methodological problem raised by the Continuum Hypothesis: in the words of Maddy, 'should the work of Gödel and Cohen be regarded as settling the continuum hypothesis, or does a mathematical question remain, amenable to solution by mathematical methods?'

Maddy's chapter shows how methodological questions within mathematics lead quickly to philosophical debates; she describes some of the philosophical responses that have been made, and concentrates upon 'naturalism'.

Let us briefly report on how these 'undecidable' statements are being viewed by working mathematicians. First we should say that, whilst a minority of practising mathematicians are seriously and thoughtfully interested in such matters, and a few become involved in polemical arguments, a substantial majority is essentially indifferent to such philosophical discussions, and has not so far felt that their daily mathematics is affected by such matters; this view may well change in the future, when the fact that different views on the size of the continuum can lead to different results in branches of mathematics apparently far removed from set theory becomes more widely known.

Second, let us examine briefly the current state of opinion on the fundamental undecidable statements that we have mentioned.

The Axiom of Choice (AC) was very controversial in the early years of this century. However, as we have explained, it is now almost universally accepted. This is surely because it has as a consequence many results that mathematicians wish to be true (a list of some of these consequences is given in the chapter of Dales, pp. 181–200); these consequences are not provable in the system of ZF.[52] It seems very unlikely that this view of AC will change significantly in the foreseeable future. Earlier in the century, the fact that the proof of a theorem depended on AC was often noted, but this practice has now almost disappeared. It is still the case that the fact that a result depends on the continuum hypothesis (CH) is usually explicitly noted. Since CH tells us the size of the continuum, it is obvious that a decision on the value of $2^{\aleph_0}$ will have a profound impact on set theory, infinite combinatorics, topology, and measure and category theories.[53] However, it was a surprise to working mathematicians that the answer to a problem in abstract analysis on the continuity of homomorphisms from the Banach algebra $C(\mathbb{I})$ of all continuous functions on the closed unit interval $\mathbb{I}$ was found to be independent of ZFC and to depend on the size of $2^{\aleph_0}$. [This example is discussed in the chapter of Dales, pp. 181–200.] There is now a substantial list of results—in areas apparently far away from set theory—which are known to depend on the size of the continuum. So what view will mathematicians take on CH in the future? This is hard to predict. At present, there may be a slight inclination among set theorists to concentrate on exploring the universe of ZFC + $\neg$CH, whilst mathematicians who are not set theorists are inclined to use CH freely when they need it to make progress. But there is certainly no consensus on whether or not CH should be accepted. Our view is that a style is evolving in which one states one's assumption on the size of the continuum, but is neutral about its preferred value.

Perhaps most interesting are the questions that are independent even of ZFC + GCH. These questions involve such famous ones as those about measurable cardinals, the Souslin hypothesis, and the Whitehead problem, which are described in the notes.[54] The resolution of such questions often involves 'large cardinal axioms'. Here is an apparently totally elementary question. Let f and g be continuous functions on the real line $\mathbb{R}$. Is $f(\mathbb{R}\setminus g(B))$ Lebesgue measurable for every Borel subset B of $\mathbb{R}$? This cannot be decided in ZFC, but is resolved positively by the assumption that there is a certain large cardinal, a *measurable cardinal.* There is an ongoing debate about which large cardinal axioms should be accepted; the basis on which such a discussion could be resolved is discussed in several chapters.

Given that, for suitable formal systems T, there exist undecidable statements φ such that there is no proof of either φ or $\neg\varphi$ from T, it is natural to consider what statements can be decided, and whether there is a natural hierarchy of such statements. One such class of statements consists of those computed by an algorithm. A historically and philosophically important doctrine is the *Church–Turing thesis*: the collection of computable, or effectively calculable, functions is exactly the collection of recursive functions. This thesis is now very generally accepted. [A full explanation of what a recursive function is is given in the chapter of Slaman, pp. 233–251. Note that the projective sets, which play a significant rôle here, appear in the chapters of both Slaman and Martin.]

Let us now return to our first question: whether or not ZF is consistent. The second incompleteness theorem of Gödel states that, if T is a theory containing arithmetic, then $T \nvdash \mathrm{Con}\, T$. Thus Con ZFC can only be proved in ZFC if ZFC is inconsistent. How then can we determine whether or not ZFC is consistent? Perhaps the main reason for mathematicians' confidence in the consistency of ZFC is the fact that, despite enormous (implicit) testing of the axioms throughout this century, no contradiction has been revealed, and so almost everyone is essentially convinced that no contradiction will appear; or, at least, if a contradiction does emerge, a modest fine-tuning of the axioms will produce a modified system that does not obviously imply any inconsistency. Is this a reasonable view? According to the view of (Bourbaki 1949):

> There are now twenty-five centuries during which mathematicians have had the practice of correcting their errors and thereby seeing their science enriched, not impoverished; this gives them the right to view the future with serenity.

[The question of the consistency of theories is the topic of the thought-provoking and somewhat uncomfortable chapter of Woodin.]

It is now time to move on, in our final section, to a brief discussion of mathematical knowledge.

7 The realism/anti-realism debate and the question about mathematical knowledge

Although there may well be those for whom, with Heraclitus:[55]

πόλεμος πάντων μὲν πατήρ ἐστι , πάντων δὲ βασιλεύς ,

it is certainly the case that strife, in a philosophical context, has often fathered unhelpful comments and positions. An example of a less than happy outcome of philosophical strife is the following *polemical remark* of A. Tarski:[56]

> ...in no interpretation of the term 'metaphysical' which is familiar and more or less intelligible to me does semantics involve any metaphysical elements peculiar to itself.

Those acquainted with the history of science are aware of the fact that many terms belonging to the language of the natural sciences change in meaning over time, often as a consequence of the change of theories. A very well-known example of this phenomenon is that provided by the term 'atom', which was originally intended to refer to indivisible components of matter.[57] However, from the time of Leucippus and Democritus the meaning of 'atom' has changed considerably, so much so that present-day physics uses it to refer to components of matter which are indeed divisible.

We find a similar situation obtaining in mathematics; terms such as 'axiom'[58] and 'function' have all been subject to change in meaning from the time they were first introduced; and given that this is a destiny that befalls most of the terms belonging to the language of all the natural sciences, it will not come as a surprise that the term 'metaphysics' has also undergone several shifts in meaning during the long and honoured service it has paid to philosophy. It is this consideration that, in our attempt to assess Tarski's claim, prompts us to take as the meaning of 'metaphysics' that currently adopted in analytical philosophy.

Problems which, within this school of thought, are classified as *metaphysical* are, as Dummett says, those:[59]

> ...about whether or not we should take a realist attitude to this or that class of entity. In any one instance, realism is a definite doctrine. Its denial, by contrast may take any one of numerous possible forms, each of which is a variety of anti-realism concerning the given subject matter: the colourless term 'anti-realism' is apt as a signal that it denotes not a specific philosophical doctrine but the rejection of a doctrine.

Having hinted at some of the problems which, within analytical philosophy, are called 'metaphysical', let us attempt to bring out the close-knit relationship between such problems and mathematical truth.

If, with the classical mathematician, we hold that to believe in the truth of $P \vee Q$ it is sufficient to show that it is impossible that both P and Q are false, we commit ourselves to the belief in the existence of a reality that the statements P and Q are about. To see this, let us assume that it has been shown that P and Q cannot be both false and that we have proof neither of P nor of Q. Clearly, in this situation, only if we believe in the existence of a reality that P and Q are about would it make sense to hold that either P or Q is true (of such a reality) and, therefore, assert $P \vee Q$. Hence, the realist's inclinations which are manifested by classical mathematicians in the philosophy of mathematics.

On the other hand, if we believe that the statement $P \vee Q$ is true just in case we can give a constructive proof either of P or of Q, then, if we have no proof either of P or of Q (or of their negations), the statement $P \vee Q$ lacks a truth value. The fact that $P \vee Q$ has no truth value dispenses with a commitment to the belief in the existence of a reality that $P \vee Q$ (and in particular P and Q) is (are) about. These considerations justify us in saying that the constructivist mathematician adopts an anti-realist position in the philosophy of mathematics. So, just as the apparently arcane *filioque* dispute—one of the causes which led to the separation of Eastern and Western Christendom—was really an outward sign of deep differences on the nature of the Trinity and of God, so the use or prohibition of the law of the excluded middle is a sign of deep differences in our perception of mathematical reality and the nature of truth in mathematics.

[Jones, in his chapter, pp. 203–214, adopts an original and interesting realist position, arising from his own experience of creating stunning new mathematics, that proof is a necessary, but not sufficient, condition for mathematical truth.]

What we have said so far in this section ought to show that *there is* a metaphysical element peculiar to semantics and, in particular, to the semantic conception of truth of mathematical statements, a metaphysical element represented by the postulation of a reality, whatever this might turn out to be, that mathematical statements are true of.

Another important philosophical problem related to the debate concerning how best we ought to understand the claim that a mathematical statement is true is that concerning mathematical knowledge. If the best characterization of a true mathematical statement is that provided by CLASS, then we are indeed right in believing that mathematics produces knowledge. This is knowledge of the reality that mathematical statements are about.

On the other hand, if the best characterization of a true mathematical statement is that offered by constructivism, the belief in the existence of mathematical knowledge would not be warranted. The reason for this is that, for the constructivist, statements and constructions, although non-arbitrarily generated, but bound by a set of conventions, are nevertheless the outcome of a process of *invention* rather than one of *discovery*.

It is our hope, in concluding this introduction, that what we have said so far might be sufficient to justify the philosophical and mathematical importance of clarifying what it means to say that a mathematical statement is true; and that the essays which follow may shed some light on this problem.

Notes

1. There are many books on the history and philosophy of Greek mathematics. See, for example, (Kline 1972), for a clear summary.

2. (Kant 1990, Introduction, §IV, p. 48):

> In all judgements in which the relation of a subject to the predicate is thought (I take into consideration affirmative judgements only, the subsequent application to negative judgements being easily made), this relation is possible in two different ways. Either the predicate B belongs to the subject A, as something which is (covertly) contained in this concept A; or B lies outside the concept A, although it does indeed stand in connection with it. In the one case I entitle the judgement analytic, in the other synthetic. Analytic judgements (affirmative) are therefore those in which the connection of the predicate with the subject is thought through identity; those in which this connection is thought without identity should be entitled synthetic. The former, as adding nothing through the predicate to the concept of the subject, but merely breaking it up into those constituents concepts that have all along been thought in it, although confusedly, can

> also be entitled explicative. The latter, on the other hand, add to the concept of the subject a predicate which has not been in any wise thought in it, and which no analysis could possibly extract from it; and they may therefore be entitled ampliative.

3. The following passage shows an application of the *a priori*/*a posteriori* distinction to the concept of knowledge. The generalization of this distinction to judgements (statements) is trivial. (Kant 1990, Introduction, §I, p. 43):

> ... we shall understand by *a priori* knowledge, not knowledge independent of this or that experience, but knowledge absolutely independent of all experience. Opposed to it is empirical knowledge, which is knowledge possible only *a posteriori*, that is, through experience.

4. In the *Critique of Pure Reason*, Kant gives the example of the arithmetical statement '$7+5=12$'; his discussion, from (Kant 1990, pp. 52–3), is as follows:

> First of all, it has to be noted that mathematical propositions, strictly so called, are always judgements *a priori*, not empirical; because they carry with them necessity, which cannot be derived from experience. If this be demurred to, I am willing to limit my statement to *pure* mathematics, the very concept of which implies that it does not contain empirical, but only pure *a priori* knowledge. We might, indeed, at first suppose that the proposition $7+5=12$ is a merely analytic proposition, and follows by the principle of contradiction from the concept of a sum of 7 and 5. But if we look more closely we find that the concept of the sum of 7 and 5 contains nothing save the union of the two numbers into one, and in this no thought is being taken as to what that single number may be which combines both. The concept of 12 is by no means already thought in merely thinking this union of 7 and 5; and I may analyse my concept of such a possible sum as long as I please, still I shall never find the 12 in it. We have to go outside these concepts, and call in the aid of the intuition which corresponds to one of them, our five fingers, for instance, or, as Segner does in his *Arithmetic*, five points, adding to the concept of 7, unit by unit, the five given in intuition. For starting with the number 7, and for the concept of 5 calling in the aid of the fingers of my hand as intuition, I now add one by one to the number 7 the units which I previously took together to form the number 5, and with the aid of that figure [the hand] see the number 12 come into being. That 5 should be added to 7, I have indeed already thought in the concept of a sum $=7+5$, but not that this sum is equivalent to the number 12. Arithmetical propositions are therefore always synthetic. This is still more evident if we take larger numbers. For it is then obvious that, however we might turn and twist our concepts, we could never, by the mere analysis of them, and without the aid of intuition, discover what [the number is that] is the sum.

5. The study of what is, for Kant, the relation between mathematics and what we have called 'pre-reflective system of representation' leads to very involved, technical discussions which would be inappropriate to even attempt to summarize here. The interested reader can consult (Friedman 1992, Part One, Ch. 2, pp. 96–135).

6. (Kant 1990, Preface to Second Edition, p. 19):

> The true method [discovered by the founder father of mathematics] was not to inspect what he discerned either in the figure, or in the bare concept of it, and from this, as it were, to read off its properties; but to bring out what was necessarily implied in the concepts that he had himself formed *a priori*, and had put into the figure in the construction by which he presented it to himself. If he is to know anything with *a priori* certainty he must not ascribe to the figure anything save what necessarily follows from what he has himself set into it in accordance with his concept.

7. (Kant 1990, Transcendental Analytic, Analytic of Principles, Ch. III, pp. 259–60):

> Take ... the concepts of mathematics, considering them first of all in their pure intuitions. Space has three dimensions; between two points there can be only one straight line, etc. Although all these principles, and the representation of the object with which this science occupies itself, are generated in the mind completely *a priori*, they would mean nothing, were we not always able to present their meaning in appearances, that is, in empirical objects. We therefore demand that a bare concept be *made sensible*, that is, that an object corresponding to it be presented in intuition. Otherwise the concept would, as we say, be without *sense*, that is, without meaning. The mathematician meets this demand by the construction of a figure, which, although produced *a priori*, is an appearance present to the senses. In the same science the concept of magnitude seeks its support and sensible meaning in number, and this in turn in the fingers, in the beads of the abacus, or in strokes and points which can be placed before the eyes. The concept itself is always *a priori* in origin, and so likewise are the synthetic principles or formulas derived from such concepts; but their employment and their relation to their professed objects can in the end be sought nowhere but in experience, of whose possibility they contain the formal conditions.

8. (Kant 1990, Preface to Second Edition, p. 18):

> The sphere of logic is quite precisely delimited; its sole concern is to give an exhaustive exposition and a strict proof of the formal rules of all thought, whether it be *a priori* or empirical, whatever be its origin or its object, and whatever hindrances, accidental or natural, it may encounter in our minds.

9. (Kant 1990, Transcendental Doctrine of Elements, Transcendental Logic, §III, The Division of General Logic into Analytic and Dialectic, p. 98).

10. (Kant 1990, Preface to Second Edition, p. 18).

11. The following quotations, from (Frege 1977, pp. 1–2), provide evidence for this thesis endorsed by Frege introducing, at the same time, the well-known distinction between the laws of thought—a thought is, for Frege, the content of a proposition, and it is the only thing concerning which the question of truth can arise—and the laws of thinking (the psychological process of ...):

> From the laws of truth there follow prescriptions about asserting, thinking, judging, inferring. And we may well speak of laws of thought in this way too. But there is at once a danger here of confusing different things. People may very well interpret the expression 'law of thought' by analogy with 'law of nature' and then have in mind general features of thinking as a mental occurrence ... In order to avoid any misunderstanding and prevent the blurring of the boundary between psychology and logic, I assign to logic the task of discovering the laws of truth, not the laws of taking things to be true or of thinking.

12. A very interesting discussion of Frege's notion of analyticity and of its relevance can be found in (Dummett 1995, Ch. 3, pp. 23–35).

13. (Frege 1884, §3, p. 4):

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

14. (Frege 1884, §3, p. 3).

15. For Frege, contrary to Kant, numbers are not objects given in the intuition (sensibility). (Frege 1950, §89, p. 101):

> I must also protest against the generality of Kant's dictum: without sensibility no object would be given to us. Nought and one are objects which cannot be given to us in sensation. And even those who hold

> that the smaller numbers are intuitable, must at least concede that they cannot be given in intuition any of the numbers greater than $1000^{1000^{1000}}$, about which nevertheless we have plenty of information. Perhaps Kant used the word 'object' in a rather different sense; but in that case he omits altogether to allow for nought or one, or for our ∞_1, — for these are not concepts either, and even of a concept Kant requires that we should attach its object to it in intuition.

Those interested in finding out more about Frege's conception of numbers as objects might benefit from reading Wright (1983) and (Dummett 1995, Chapter 11, pp. 131–40, and Chapter 18, pp. 223–40).

16. (Marion 1995, p. 191).

17. An example of a definition satisfying Kronecker's requirements is that of prime number:

Definition *An integer $p > 1$ is a* prime *if its only divisors are* 1 *and p. An integer greater than* 1 *which is not a prime is termed* composite.

We have a simple procedure to decide whether an integer k is prime or composite: we check to see whether k is divisible by any of the positive integers d such that $1 < d < k$. If we find at least one such d then k is composite; otherwise k is prime. Notice that since there are finitely many values of d such that $1 < d < k$, and we can decide in a finite number of steps whether or not a positive integer k is divisible by a positive integer d, our procedure of decision will have to halt after a finite number of steps and produce an answer, and, therefore, it deserves to be called an 'algorithm'.

18. An example of a theorem satisfying Kronecker's requirements is the following:

Theorem *Every statement form $\mathcal{A}$ is logically equivalent to one in disjunctive normal form.*

Proof Given a statement form $\mathcal{A}$, compute its truth table. Isolate the rows of the truth table where $\mathcal{A}$ takes the value 'true'; say, $\mathcal{A}$ takes the value **T** only in rows $r_1, \dots, r_n$. If r_i is a row such that $1 \leq i \leq n$, examine which truth value the corresponding structure assigns to the variables occurring in $\mathcal{A}$, and, if a variable A has been assigned the truth value **T**, write down A; if it has been assigned the truth value **F**, write down $\neg A$. When you have completed this procedure for all the variables occurring in $\mathcal{A}$ in relation to r_i, form the conjunction Γ_i of the formulae you have written down. If $\mathcal{A}$ contains m different variables, then $\Gamma_i = \bigwedge_{k=1}^{m} A_k^{'}$, where $A_k^{'} = A_k$ if A_k is assigned the value **T** in r_i; $A_k^{'} = \neg A_k$, if A_k is assigned the value **F** in r_i. Once you have completed this routine for all the rows $r_1, \dots, r_n$, form the disjunction $\mathcal{A}^* = \bigvee_{i=1}^{n} \Gamma_i$. The statement form $\mathcal{A}^*$ will be logically equivalent to $\mathcal{A}$. Since a statement form contains a finite number of variables and connectives the procedure must be completed in a finite number of steps and, therefore, is an algorithm. □

19. A classical example of *logical proof* is that of the following theorem.

Theorem *Suppose that a and b are any positive integers. Then there exists a positive integer n such that $na \geq b$.*

Proof Assume that the statement of the theorem is not true, so that for some a and b, $na < b$ for every positive integer n. Then the set

$$S = \{b - na : n \text{ is a positive integer}\}$$

consists entirely of positive integers. By the well-ordering theorem, S will possess a least element, say $b - ma$. But $b - (m+1)a$ also lies in S, since S contains all integers of this form. Furthermore, we have

$$b - (m+1)a = (b - ma) - a < b - ma,$$

contrary to the choice of $b - ma$ as the smallest integer in S. □

The above theorem and proof have been taken almost word for word from (Burton 1980, Ch. 1, §1.1, pp. 2–3). Notice that the proof given does not provide a finite procedure such that, given any two positive integers a and b, we can find a positive integer n such that $na \geq b$. Of course, the trivial proof of the theorem ('Take $n = b$') does provide such an n.

A second logical proof establishes a result, the *intermediate value theorem* for continuous functions, that is the basic foundation of an enormous amount of analysis.

Theorem *Let f be a continuous function on the closed interval $[a, b]$. Suppose that $f(a) < 0 < f(b)$. Then there exists x in (a, b) such that $f(x) = 0$.* □

The proof of the theorem is immediate from the fact that a non-empty set of real numbers that is bounded above has a supremum. The standard proof is a logical proof because it uses proof by contradiction and does not give an algorithm which produces the number x. What is the status of the fact that a set bounded above has a supremum? It could be an axiom, the axiom that states that an ordered field is (Dedekind) complete. [See the chapter of Dales, pp. 181–200.] It could be an attempt to describe the 'reality' of the real line.

20. Let us take as an example von Neumann's universe V. The class V is a typed universe with the following defining properties: i) $V_0 = \emptyset$; ii) $V_{\alpha+1} = \mathcal{P}(V_\alpha)$ if α is not a limit ordinal; iii) $V_\alpha = \bigcup_{\beta<\alpha} V_\beta$ if α is a limit ordinal. We call **set** any of the totalities occurring in V. A fundamental characteristic of V is that the *first* occurrence of a totality X in V is at a level which is strictly above the level of the *first* occurrence of any of the elements of X. Keeping this in mind, we can now prove that the totalities $A = \{x : x \text{ is a } \mathbf{set}\}$, $R = \{y : y \text{ is a } \mathbf{set} \wedge \neg(y \in y)\}$ and $BF = \{z : z \text{ is an ordinal}\}$ are not **sets**. In the case of A there is no type V_α in which A can occur because any level of V contains **sets** which occur there for the first time. Secondly, since for any X occurring in V we have that $\neg(X \in X)$, if X occurs in V then $X \in R$ and we can apply the same reasoning as above to show that R cannot occur in V. Lastly, if we choose the von Neumann definition of ordinal number and his representation of finite ordinals, that is, $0 \mapsto \emptyset, \ldots, n' \mapsto n \cup \{n\}, \ldots$, we can conclude that, since each level of V contains

the first occurrence of an ordinal, BF does not occur in V. The totalities A, R and BF are *too large* to be **sets**.

21 For a stimulating discussion of Hilbert's programme see (Detlefsen 1986).

22. Things are often not so simple as they might seem to be at first sight. For quite some time there has been a lively discussion about what, according to Hilbert, the finitary part of number theory might be. Most of the scholars seem now to agree that this is embodied in what is known as Primitive Recursive Arithmetic—a rigorous characterization of Primitive Recursive Arithmetic can be found in (Troelstra and van Dalen 1988, Chapter 3, §2, pp. 120–6). On these issues see (Tait 1981) and (Detlefsen 1986, Chapter 1, §2, and Chapter 2, §2).

23. If we want to be very rigorous, we ought to say that, for Hilbert, the numerals '355' and '356' are simply names for totalities whose elements are '1s', that is, in the same way that '2' is the name for the totality 11 and '3' the name for the totality 111, so '355' and '356' are the names for the appropriate totalities (of 1s).

24. (Hilbert 1983, p. 195):

> Let us remember that *we are mathematicians* and that as mathematicians we have often been in precarious situations from which we have been rescued by the ingenious method of ideal elements. I showed you some illustrious examples of the use of this method at the beginning of this chapter. Just as i = $\sqrt{-1}$ was introduced to preserve in the simplest form the laws of algebra (for example, the laws about the existence and number of roots of an equation); just as ideal factors were introduced to preserve the simple laws of divisibility for algebraic whole numbers (for example, a common ideal divisor for the numbers 2 and $1 + \sqrt{-5}$ was introduced, though no such divisor really exists); similarly, to preserve the simple formal rules of ordinary Aristotelian logic, we must *supplement the finitary statements with ideal statements.*

This particular stand on the value of ideal statements is what has motivated some authors in classifying Hilbert as an instrumentalist. We disagree with this view because Hilbert's instrumentalism seems to be localized exclusively in the area of ideal statements.

25. (Hilbert 1983, p. 199):

> There is just one condition, albeit an absolutely necessary one, connected with the method of ideal elements. That condition is a *proof of consistency*, for the extension of a domain by the addition of ideal elements is legitimate only if the extension does not cause contradictions to appear in the old, narrower domain, or, in other words, only if the relations that obtain among the old structures when the ideal structures are deleted are always valid in the old domain.

26. See (Detlefsen 1986, Chapters III–V).

27. See (Kleene 1952, Part II, Chapter VIII, §42, Theorem 30, p. 210); and (Mendelson 1987, Chapter III, §5, p. 166). Concerning the so-called 'partial realizations' of Hilbert's programme, see (Simpson 1988). More information about Hilbert's programme may be found in (Sieg 1988) and in (Feferman 1988).

28. (Hilbert 1983, pp. 191–2):

> Material logical deduction is indespensable. It deceives us only when we form arbitrary abstract definitions, especially those which involve infinitely many objects. In such cases we have illegitimately used material logical deduction; i.e., we have not paid sufficient attention to the preconditions necessary for its valid use. In recognising that there are such preconditions that must be taken into account, we find ourselves in agreement with the philosophers, notably with Kant. Kant taught — and it is an integral part of his doctrine — that mathematics treats a subject matter which is given independently of logic. Mathematics, therefore, can never be grounded solely on logic. Consequently, Frege's and Dedekind's attempts to so ground it were doomed to failure. As a further precondition for using logical deduction and carrying out logical operations, something must be given in conception, viz., certain extralogical concrete objects which are intuited as directly experienced prior to all thinking. For logical deduction to be certain, we must be able to see every aspect of these objects, and their properties, differences, sequences, and contiguities must be given, together with the objects themselves, as something which cannot be reduced to something else and which requires no reducction. This is the basic philosophy which I find necessary, not just for mathematics, but for all scientific thinking, understanding, and communicating. The subject matter of mathematics is, in accordance with this theory, the concrete symbols themselves whose structure is immediately clear and recognizable.

29. (Heyting 1971, §2.2.2):

> Every mathematical assertion can be expressed in the form: 'I have effected the construction A in my mind'. The mathematical negation of this assertion can be expressed as 'I have effected in my mind a construction B, which deduces a contradiction from the supposition that the construction A were brought to an end', which is again of the same form.

30. (van Stigt 1990, Chapter IV, §4.5.2, p. 167):

> The lack of simplicity and uniformity in the domain of mathematical constructions is mainly due to the freedom of the Subject to create ever more and more complex constructions, using Intuition and previously constructed entities and tools in 'a free unfolding'. The

> condition 'previously constructed' is not just a restriction, it is a sanction of their mathematical birth-right to be exploited to the full. For example, (Brouwer 1913, p. 81) points out that in the construction of the infinite ordinal ω 'of course, every previously constructed set of every previously performed constructive operation may serve as the unit.' In this free-unfolding, however, the 'previously acquired' can only be a general principle for a genetic hierarchy of complex constructions, and was used as such in e.g. Brouwer's hierarchy of species. A comprehensive hierarchical classification of all mathematical constructions must remain illusory.

Since the time of Brouwer, Intuitionism has developed into a broad church with different people holding different opinions on certain things. The main differences of opinion are on mathematics as a construction of the thought and on the importance of language in mathematics. See (Dummett 1985).

31. (Brouwer 1913, p. 82):

> In the domain of finite sets in which the formalist axioms have an interpretation perfectly clear to the intuitionists, unreservedly agreed to by them, the two tendencies differ solely in their method, not in their results; this becomes quite different however in the domain of infinite or transfinite sets, where, mainly by the application of the axiom of inclusion, quoted above, the formalist introduces various concepts, entirely meaningless to the intuitionist, such as for instance '*the set whose elements are the points of space,*' '*the set whose elements are the continuous functions of a variable,*' '*the set whose elements are the discontinuous functions of a variable,*' and so forth.

32. For more details, see (Dummett 1977, Chapter 3, §3.1, pp. 55–65).

33. (Dummett 1977, p. 11).

34. It is important to notice that classical logic and arithmetic are very similar to intuitionistic logic and arithmetic, as has been proved by Gödel and others; see (Dummett 1977, Ch. 2, §2.1, p. 36) and (Troelstra and van Dalen 1988, Ch. 2, §3, pp. 57–59). The radical differences between the two approaches begin to emerge in analysis.

35. See (Tarski 1952) and (Tarski 1956).

36. The debate about whether considerations similar to those relative to formalized languages are also applicable to ordinary language are beyond the scope of the present discussion. We are here concerned only with truth in mathematics. However, those interested in these other aspects of the problem can see (Kripke 1975) and (Kirkham 1995, Ch. 9).

37. (Tarski 1952, §3, p. 15).

38. Remember that, for Tarski, 1) can be considered only as a *partial* definition of truth in the sense that it explains on what the truth of the sentence:

'the cat is on the mat', rests; and that, if we substitute in 1) the sentence variable p for the sentence: 'the cat is on the mat'—on the right-hand side of the biconditional—and the term X for 'the cat is on the mat'—on the left-hand side of the biconditional—what we obtain:

1*) X is true if and only if p,

cannot be a definition of truth either, because 1*) is not a sentence, but a schema known as T-schema.

39. A language L is *semantically closed* just in case: i) L containes the names of its expressions; ii) L contains the predicate *true*, and such a predicate can be applied to sentences of L; iii) all the sentences which determine the use of *true* are expressible in L.

40. The essential conditions which, if satisfied, would dispense with the liar paradox are: that the definition of truth relating to sentences of L and all the equivalences of the form:

2*) X is true if and only if p,

where p belongs to L, can only be formulated in the meta-language $M(L)$ of L; and that $M(L)$ must be *essentially richer* than L. The reasons for this are that: i) the sentences about sentences of L which could generate the paradox would not belong to L, but to $M(L)$ and, ii) if $M(L)$ is *essentially richer* than L, this implies that there is no way of representing (translating) $M(L)$ into L.

41. The system of abbreviations CLASS, INT, RUSS and BISH is taken from Bridges's chapter in this volume.

42. This definition is taken from (Mendelson 1987, Chapter 2, §2, p. 48).

43. Here Σ is the set of all denumerable sequences (as opposed to finite sequences) of elements of the domain D of the interpretation $\mathfrak{M}$. For the definition of the notion of *satisfaction*, see (Mendelson 1987, Chapter 2, §2, p. 48).

44. See (Mendelson 1987, ibid. p. 49).

45. The CLASS interpretation of the logical constants is the typical model-theoretical interpretation, which answers questions of the type 'When is it true to assert $\neg P$?', etc., by producing truth-tables, see (Mendelson 1987, Chapter 1, §1); and questions of the type 'When is it true to assert $\forall x F(x)$?', etc., through the standard use of the Tarskian concept of satisfiability (ibid., Chapter 2, §2). The BHK interpretation of the logical constants is given in terms of *provability*, namely, it is aimed at answering questions of the type 'What is a proof of $\neg P$?', etc. We shall hereby quote from (Troelstra and van Dalen 1988) what are called *Heyting's axioms*:

H1 A proof of $A \wedge B$ is given by presenting a proof of A and a proof of B.

H2 A proof of $A \vee B$ is given by presenting either a proof of A or a proof of B (plus the stipulation that we want to regard the proof presented as evidence for $A \vee B$).

H3 A proof of $A \rightarrow B$ is a construction which permits us to transform any proof of A into a proof of B.

H4 Absurdity $\perp$ (contradiction) has no proof; a proof of $\neg A$ is a construction which transforms any hypothetical proof of A into a proof of a contradiction.

H5 A proof of $\forall x A(x)$ is a construction which transforms a proof of $d \in D$ (D the intended range of the variable x) into a proof of $A(d)$.

H6 A proof of $\exists x A(x)$ is given by providing a $d \in D$, and a proof of $A(d)$.

Heyting's axioms were originally produced by A. Heyting to axiomatize intuitionistic logic, but, in fact, play a much wider rôle in constructivism, because 'The constructivist mathematician interprets connectives and quantifiers according to intuitionistic logic.' (Bridges and Richman 1988, Chapter 1, §3, p. 10).

46. If $P \vee \neg P$ were a law of constructive logic, asserting it would imply, according to axiom $H2$ of the BHK interpretation, that we have either a proof of P or a proof of $\neg P$, for any proposition P. But if P is, for instance, a mathematical conjecture which has not yet been proved (or refuted), we would not be able to assert $P \vee \neg P$. Note how different the CLASS interpretation of $P \vee \neg P$ is from the BHK interpretation. Of course, $P \vee \neg P$ is a law of CLASS logic.

47. There are numerous accounts of Gödel's completeness theorem. See, for example (Barwise 1977, Chang and Keisler 1973, Enderton 1972, and Mendelson 1987).

48. For a study of the mathematics and philosophy of Cantor, see (Dauben 1979).

49. For an historical account of the controversy about the Axiom of Choice, see (Moore 1982); for a mathematical discussion of the significance of this axiom, see (Jech 1973).

50. For Gödel's results, see (Gödel 1931). For a discussion of the incompleteness theorems, see (Smorynski 1977); the results are proved in many standard texts, such as (Enderton 1972).

51. Cohen's original paper was (Cohen 1963–64); see also (Cohen 1966). There are now many methods in the standard texts of proving that CH is independent of ZFC, and other independence results; the methods are surveyed in (Kunen 1980, VII). See, for example, (Jech 1978). A rather elementary approach, designed to appeal to non-logicians, using Boolean-valued models, is given in (Dales and Woodin 1987).

52. There is a weaker version of AC, called DC, the *Axiom of Dependent Choice*. This allows some results to be proved, but avoids some 'undesirable' consequences of the full AC. It is known that, under a certain large cardinal axiom, there are models of ZF + DC in which AC is false; in these models, every subset of $\mathbb{R}$ is Lebesgue measurable and every linear map from a Banach space into another Banach space is continuous. These might be thought to be desirable results. However, few mathematicians have chosen to work in this setting; they prefer the full power of AC, at the cost of non-measurable subsets of $\mathbb{R}$ and discontinuous linear operators.

53. Here is a sample application in set theory itself. We write MA for *Martin's Axiom*: this is an axiom introduced by Martin and Solovay in (1970) to settle questions in ZFC + ¬CH that could not otherwise be resolved. The cardinal numbers $2^{\aleph_0}$ and $2^{\aleph_1}$ are clearly basic in any theory of the 'size' of infinite sets, and certainly $2^{\aleph_1} \geq 2^{\aleph_0}$. However, whether or not $2^{\aleph_1} = 2^{\aleph_0}$ cannot be resolved in ZFC: with CH, $2^{\aleph_1} > 2^{\aleph_0}$, but, with MA + ¬CH, $2^{\aleph_1} = 2^{\aleph_0}$.

54. We describe briefly these three famous questions. We must utilize some standard mathematical terminology without explanation; we hope that the reader will at least absorb something of the flavour of these questions.

(i) A $\{0, 1\}$-*valued measure* on a non-empty set S is a countably additive function defined on the family of all subsets of S and assuming only the values 0 and 1. A cardinal number κ is *measurable* if there is such a measure μ on κ with $\mu(\kappa) = 1$ and $\mu(\{x\}) = 0$ for each $x \in \kappa$. The class of non-measurable cardinals is very extensive; it contains $\aleph_0$ and is closed under all the standard operations of cardinal arithmetic. It cannot be proved in ZFC that measurable cardinals exist; the axiom that there is such a cardinal is a *large cardinal axiom.*

There is now a multitude of 'large cardinal axioms', and an industry that describes the relationships between these axioms. What is striking is that the axioms so far promulgated seem to fit into a beautiful pattern.

(ii) Consider the real line $\mathbb{R}$ with the usual order $\leq$. Then $(\mathbb{R}, \leq)$ is a totally ordered set with neither a maximum nor a minimum element, and $\mathbb{R}$ is connected and separable with respect to the order topology, which is the usual topology on $\mathbb{R}$. It is easy to see that these properties characterize $(\mathbb{R}, \leq)$, in the sense that any totally ordered set $(S, \leq)$ with these properties is order-isomorphic to $(\mathbb{R}, \leq)$. In 1920, Souslin raised the question whether 'separable' could be replaced by 'every collection of pairwise disjoint open intervals in $(S, \leq)$ is countable'. The *Souslin hypothesis* (SH) is that every totally ordered set satisfying this latter condition is separable, and hence order-isomorphic to $(\mathbb{R}, \leq)$. Much later than 1920, it was proved that this hypothesis is independent of ZFC + CH: SH follows from MA + ¬CH, and, by a result of Jensen, SH is consistent with GCH, but ¬SH follows from a principle $\diamondsuit$, called 'diamond', which is consistent with ZFC + GCH. See Jech (1978) and Kunen (1980), for example.

(iii) We now give an example from algebra.

Let $(G, +)$ be an abelian group. A subset S of G is *linearly independent* if $n_1 = \cdots = n_k = 0$ whenever $n_1 s_1 + \cdots + n_k s_k = 0$ for $n_1, \ldots, n_k \in \mathbb{Z}$ and distinct elements $s_1, \ldots, s_k$ of S, and S *spans* G if each element of G can be written as $n_1 s_1 + \cdots + n_k s_k$ for some $n_1, \ldots, n_k \in \mathbb{Z}$ and $s_1, \ldots, s_k \in S$. The group G is *free* if it has a linearly independent subset that spans G. There is a clear sense in which the free groups are the basic groups from which other groups can be obtained. An abelian group G is a *Whitehead group* if, whenever $\pi : G \to H$ is a surjective group homomorphism from G onto another group H such that the kernel of π is isomorphic to $\mathbb{Z}$, there is a group homomorphism $\rho : H \to G$ such that $\pi(\rho(t)) = t \ (t \in H)$. (The latter condition says that π *splits.*) It is easy to see that a free group is a Whitehead group. It was a fundamental problem of Whitehead whether the converse holds: is every Whitehead group free? This question too is undecidable in ZFC. In Gödel's constructible universe, every Whitehead group is free, but MA + ¬CH implies that there is a Whitehead group which is not free.

In fact CH does not resolve the question: there is a model of ZFC + CH in which there is a Whitehead group which is not free. See Ekloff (1977) for a discussion of this problem.

55. This is (Heraclitus, Fragment 53).

56. See (Tarski 1956, §19, p. 36).

57. The word 'atom' originates from the ancient Greek word ἄτομος, which derives from the verb τέμνω =: I cut, I divide; and the prefix ἀ− which was used to negate the concept expressed by the suffix.

58. See (Oliveri 1997*b*).

59. (Dummett 1991, Introduction, p. 4).

Bibliography

Barwise, J. (1977). An introduction to first-order logic. In *Handbook of mathematical logic* (ed. J. Barwise), pp. 5–46. North-Holland, Amsterdam.

Benacerraf, P. and Putnam, H. (ed.) (1983). *Philosophy of mathematics: selected readings* (2nd edn). Cambridge University Press.

Bourbaki, N. (1949). Foundations of mathematics for the working mathematician. *Journal of Symbolic Logic*, **14**, 1–8.

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

Bridges, D. S. and Richman, F. (1987). *Varieties of constructive mathematics.* London Math. Soc. Lecture Notes, Vol. 97. Cambridge University Press.

Brouwer, L. E. J. (1913). Intuitionism and formalism. *Bull. American Math. Soc.*, **20**, 81–96. Reprinted in *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 77–89. Cambridge University Press.

Burton, D. M. (1980). *Elementary number theory.* Allyn and Bacon, Boston, Massachusetts.

Cantor, G. (1883). *Grundlagen einer allgemeinen Mannigfaltigheitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen.* Teubner, Leipzig. English translation: Foundations of the theory of manifolds (trans. U. Parpart), *The Campaigner*, **9** (1976), 60–96.

Chang, C. C. and Keisler, H. J. (1973). *Model theory.* North-Holland, Amsterdam.

Cohen, P. J. (1963–64). The independence of the continuum hypothesis. *Proc. National Academy Sciences*, **50**, 1143–8; **51**, 105–10.

Cohen, P. J. (1966). *Set theory and the continuum hypothesis.* W. A. Benjamin, New York.

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.

Dauben, J. W. (1979). *Georg Cantor. His mathematics and philosophy of the infinite.* Harvard University Press Cambridge, Massachusetts. Reprinted in paperback by Princeton University Press, 1990.

Detlefsen, M. (1986). *Hilbert's program,* Synthèse Library, Vol. 182. Kluwer Academic, Dordrecht.

Dummett, M. A. E. (1977) *Elements of intuitionism.* Clarendon Press, Oxford.

Dummett, M. A. E. (1983). The philosophical basis of intuitionistic logic. In *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 97–129. Cambridge University Press.

Dummett, M. A. E. (1991). *The logical basis of metaphysics.* Duckworth, London.

Dummett, M. A. E. (1995). *Frege: philosophy of mathematics,* Duckworth, London.

Ekloff, P. C. (1977). Whitehead's problem is undecidable. *American Math. Monthly,* **83**, 775–88.

Enderton, H. (1972). *A mathematical introduction to logic.* Academic Press, New York.

Feferman, S. (1988). Hilbert program relativized: Proof theoretical and foundational reduction. *Journal of Symbolic Logic,* **53**, 364–84.

Frege, G. (1893, 1903). *Grundgesetze der Arithmetic, begriffsschriftlich abgeleitet, I and II.* Hermann Pohle, Jena. Translated as *The basic laws of arithmetic* (trans. M. Furth). University of California Press, Berkeley, 1964.

Frege, G. (1977). Thoughts. In *Logical investigations,* (ed. P. Geach), pp. 1–30. Blackwells, Oxford.

Frege, G. (1884). *Die Grundlagen der Arithmetik.* W. Koebner, Berlin. Translated as *The foundations of arithmetic* (2nd revised edn) (trans. J. L. Austin). Blackwells, Oxford, 1989.

Friedman, M. (1992). *Kant and the exact sciences.* Harvard University Press, Cambridge, Massachusetts.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatsh. Math. Phys.,* **38**, 173–98. Reprinted as: On formally undecidable propositions of *Principia Mathematica* and related systems I. In *Kurt Gödel: collected works,* Vol. I (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort). Oxford University Press, New York, 1986.

Heraclitus, fragment 53 DK. In G. Colli, *La sapienza Greca,* Vol. III, p. 34, Eraclito, Milano, 1980.

Heyting, A. (1962). After thirty years. In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (ed. E. Nagel, P. Suppes, and A. Tarski), pp. 194–7. Stanford University Press, California.

Heyting, A. (1971). *Intuitionism—an introduction* (3rd edn). North-Holland, Amsterdam.

Hilbert, D. (1983). On the infinite. In *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 97–129. Cambridge University Press.

Jech, T. (1973). *The axiom of choice.* North-Holland, Amsterdam.

Jech, T. (1978). *Set theory.* Academic Press, New York.

Kant, I. (1990). *Critique of pure reason* (trans. N. K. Smith). Macmillan, London.

Kirkham, R. L. (1995). *Theories of truth.* MIT Press, Cambridge, Massachusetts.

Kleene, S. C. (1952). *An introduction to metamathematics.* von Nostrand, Princeton. Reprinted by North-Holland, Amsterdam, 1974.

Kline, M. (1972). *Mathematical thought from ancient to modern times.* Oxford University Press, New York. Reprinted in paperback in three volumes, 1990.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690–716.

Kunen, K. (1980). *Set theory, an introduction to independence proofs.* North-Holland, Amsterdam.

Linsky, L. (ed). (1952). *Semantics and the philosophy of language*, The University of Illinois Press at Urbana.

Marion, M. (1995). Kronecker's 'Safe haven of real mathematics'. In *Québec studies in the philosophy of science* (ed. M. Marion and A. S. Cohen), pp. 189–215. Kluwer, Dordrecht.

Martin, D. A. and Solovay, R. M. (1970). Internal Cohen extensions. *Annals of Mathematical Logic*, **12**, 143–78.

Martin-Löf, P. (1998). Truth and knowability: on the principles C and K of Michael Dummett. *This volume*, 105–14.

Mendelson, E. (1987). *Introduction to mathematical logic.* (3rd edn.). The Wadsworth & Brooks/Cole mathematics series, Belmont, California.

Moore, G. H. (1982). *Zermelo's Axiom of Choice: its origins, development, and influence.* Springer-Verlag, New York.

Oliveri, G. (1997*b*). Criticism and growth of mathematical knowledge. *Philosophia Mathematica* (III), **5**, 228–49.

Russell, B. and Whitehead, A. N. (1910, 1912, 1913). *Principia Mathematica*, Vols. I, II, and III. Cambridge University Press.

Sieg, W. (1988). Hilbert's programme sixty years later. *Journal of Symbolic Logic*, **53**, 338–48.

Simpson, S. G. (1988). Partial realizations of Hilbert's program. *Journal of Symbolic Logic*, **53**, 349–63.

Smorynski, C. (1977). The incompleteness theorems. In *Handbook of mathematical logic* (ed. J. Barwise), pp. 821–65. North-Holland, Amsterdam.

Tait, W. (1981). Finitism. *Journal of Philosophy*, **78**, 524–46.

Tarski, A. (1952). The semantic conception of truth. In *Semantics and the philosophy of language* (ed. L. Linsky), pp. 13–47. The University of Illinois Press at Urbana.

Tarski, A. (1956). The concept of truth in formalized languages. In *Logic, semantics, metamathematics* (ed. A. Tarski) pp. 152–278. Oxford University Press.

Troelstra, A. S., and van Dalen, D. (1988). *Constructivism in mathematics. An introduction*, Vols. I and II. North-Holland, Amsterdam.

van Stigt, W. P. (1990). *Brouwer's intuitionism.* North-Holland, Amsterdam.

Wright, C. (1983). *Frege's conception of numbers as objects.* Aberdeen University Press.

Department of Pure Mathematics
University of Leeds
Leeds LS2 9JT
England
email: pmt6hgd@leeds.ac.uk

Wolfson College
Oxford OX2 6UD
England
email: gianluigi.oliveri@wolfson.ox.ac.uk

PART I

Knowability, constructivity, and truth

2

Truth and objectivity from a verificationist point of view

Dag Prawitz

Truth in mathematics is a very appropriate theme for a conference on what mathematics is about. Questions of truth often work as a catalyst in bringing out contrasts between different positions concerning the nature of a given field. For instance, the question of the applicability of truth is a main issue in discussions on the position known as *mathematical formalism*, a position that occurs in many versions. Hilbert's variant of formalism is based on the idea that so-called real sentences are true or false while all other sentences lack truth values. In his chapter in this volume, Dales (1998) seems to advocate a more radical formalism. Challenging the realist conception of mathematics by saying that it takes truth as a key notion but leaves unanswered how truth is established, he takes the merit of formalism to be that it accounts for the nature of mathematics without relying upon any notion of truth. The critics of such a strict formalism typically claim that in the end even the formalist account will depend on some notion of truth. I think that such a criticism can be rightly levelled also against Dales's formalism. He proposes that a mathematical theorem is to be understood as asserting that a certain formula follows logically from certain axioms, but if *follows logically* is understood as *follows according to the rules of predicate calculus*, then the proposal reduces mathematics to a body of truths about a certain calculus, a certain formal game if you want. This makes mathematics a science that still pursues truths, although truths of a special kind, and we are left with explaining the nature of such truths.

I am though in partial agreement with Dales's thinking that realism is unsatisfactory in not providing an informative explanation of its concept of truth. In this chapter I shall consider an alternative, constructive approach to truth. In other words, I shall discuss how truth is to be understood from the point of view of intuitionism or verificationism. I use the term *verificationism* to indicate that, following Michael Dummett, I am thinking of intuitionism as based on considerations of meaning approached from a verificationist point of view rather than on considerations of an ontological kind.

The kind of verificationism that I have been interested in is inspired by some

This chapter was written in the context of a project 'Meaning and Interpretation', partly supported by The Bank of Sweden Tercentenary Foundation.

of Gentzen's work, which may not be known to everyone. Therefore, I shall first say something about the main idea of Gentzen (1934–35) that is relevant here.

1 Gentzen's idea

One of Gentzen's main ideas, which underlies his system of natural deduction, is that the meaning of our logical concepts is determined by special kinds of inference rules that Gentzen called *introduction rules.* For instance, the introduction rules for implication, the existential quantifier, and the universal quantifier may be written respectively as follows

$$\frac{\begin{array}{c}[A]\\ \vert\\ B\end{array}}{A \to B} \qquad\qquad \frac{A(t)}{\exists x A(x)} \qquad\qquad \frac{\begin{array}{c}\vert\ (a)\\ A(a)\end{array}}{\forall x A(x)},$$

where A within square brackets indicates that assumptions of the form A occurring in the derivation of B are discharged at an inference of this kind so that the conclusion $A \to B$ becomes independent of them, and where $\vert\ (a)$ indicates that the parameter a occurring in the derivation of $A(a)$ is bound when $\forall x A(x)$ is inferred, which requires that a does not occur in an assumption that A depends on. Accordingly, Gentzen's idea is that we are to understand $A \to B$ as saying that there is a derivation of B from the assumption A, $\exists x A(x)$ as saying that for some term t there is a derivation of $A(t)$, and $\forall x A(x)$ as saying that there is a free variable derivation of $A(a)$, that is, a derivation of $A(a)$ for an arbitrary a about which no assumptions are made.

This may sound as a formalist position: the meaning of the logical constants are given by inference rules that are just taken for granted, in other words, their validity is not thought of as depending on anything. However, Gentzen's point was that only inference rules of a certain kind, that is, the introduction rules, were given as valid in the sense that they just state what we mean by the sentences that occur as conclusions of the inferences in question. Other inference rules were to be justified on the basis of the meaning given to the logical constants by the introduction rules. For instance, we can justify modus ponens, what Gentzen calls the elimination rules for →, where a formula B is inferred from premisses A and $A \to B$ giving rise to a derivation of the form

$$\frac{\begin{array}{cc}\vert & \vert\\ A & A \to B\end{array}}{B}$$

as follows: since, in view of the meaning of →, the premiss $A \to B$ guarantees that there is a derivation

$$\begin{array}{c}A\\ \vert\\ B\end{array}$$

that is, a derivation of B from A, we may in this derivation replace the assumption A by the derivation of A guaranteed by the other premiss, and we then have a derivation of B, the conclusion of the inference in question, in which B does not depend on the assumption A. Diagrammatically we may picture the given situation by the figure to the left below, where the derivation of B from A guaranteed by the premiss $A \rightarrow B$ is inserted above $A \rightarrow B$; a derivation of this kind can thus be transformed into a derivation of the form shown to the right below:

$$\dfrac{\begin{array}{c}\vert\\ A\end{array} \quad \dfrac{\begin{array}{c}[A]\\ \vert\\ B\end{array}}{A \rightarrow B}}{B} \qquad\qquad \begin{array}{c}\vert\\ A\\ \vert\\ B\end{array}$$

Similarly, universal instantiation, the elimination rule for $\forall$, where a conclusion $A(t)$ is inferred from a premiss $\forall x A(x)$, is justified as follows: given the free variable derivation of $A(a)$ guaranteed by the meaning of the premiss, we may in this derivation replace all the occurrences of the parameter a, which cannot occur in any assumption on which $A(a)$ depends, by the term t, and we then have a derivation of $A(t)$, the conclusion of the inference in question. We have here transformed a derivation of the form shown to the left below into one of the form shown to the right below:

$$\dfrac{\dfrac{\begin{array}{c}\vert\,(a)\\ A(a)\end{array}}{\forall x A(x)}}{A(t)} \qquad\qquad \begin{array}{c}\vert\,(t)\\ A(t)\,.\end{array}$$

Gentzen remarked that an introduction rule 'defines' the meaning of the logical constant exhibited in the conclusion and that an elimination rule is justified by this definition. However, being aware of the fact that the inference rules are not real definitions, he did not make much of this remark, that is, he did not try to spell it out in a more coherent way. Nevertheless, the idea hinted at in this somewhat inadequate way is clearly fundamental for his Hauptsatz, his result about cut-free proofs. Above I have explained the idea in terms of certain transformations of derivations. These transformations are identical to so-called reductions used in proof theory to normalize proofs. It can be shown (Prawitz 1965, 1971) that by carrying out such reductions we obtain a normal proof in which no formula simultaneously stands as the conclusion of an introduction inference and as major premiss of an elimination inference, a result which is equivalent to Gentzen's Hauptsatz.

Gentzen's idea, which was formulated for first-order predicate logic, has been extended to other areas. For instance, Martin-Löf (1971) showed how to extend it to first-order arithmetic, where

$$N(0) \qquad\qquad \dfrac{N(t)}{N(t')}$$

are taken as introduction rules for the assertion that something is a natural number, and the principle of induction

$$\frac{N(t) \quad A(0) \quad \begin{array}{c}A(a)\\ |\\ A(a')\end{array}}{A(t)}$$

is the corresponding elimination rule, which can be justified by the introduction rules in the same general way as explained above, essentially amounting to the kind of transformation used by Gentzen (1938) in his consistency proof for arithmetic. Martin-Löf's type theory (1974) is a further extension by which several mathematical concepts are analysed in a Gentzen-like way.

2 Basic ideas of verificationism

A way of formulating Gentzen's idea more systematically in semantic terms is to start from the idea that the meaning of a statement is determined by what counts as a proof of the statement, an idea formulated already before Gentzen by Heyting in his account of intuitionism. We can then say, for example, that the introduction rule for the existential quantifier states that an existential statement $\exists x A(x)$ is to be understood as a statement such that a proof of $A(t)$ for some term t is what counts as a proof of it—in this way it gives the meaning of $\exists x A(x)$. We see here the similarity between Gentzen's idea and the intuitionistic explanations of the logical constants; compare also what Bridges (1998) calls the 'key feature of constructive mathematics' in his chapter of this volume. We also see the importance of the fact that the sentences occurring as premisses of the introduction rules are of lower complexity than that of the conclusion: because of this feature, the statement of what counts as a proof of a sentence can be taken as a recursive clause in an explanation of the meaning of the sentence.

One must add here an important distinction between *canonical* proofs and *indirect* proofs. An introduction rule for a sentence in Gentzen's system amounts to a statement of what counts as a canonical proof of the sentence in question. Of course, the sentence can be proved also in other ways by inferring it as the conclusion of the application of an elimination rule; we then say that we have an indirect proof. The justification of the elimination rule shows how such an indirect proof can be transformed to canonical form; or, more precisely, that is what the normalization theorem says—by successive applications of the reductions (that correspond to the justifications of the elimination rules) we obtain a normal proof which is in canonical form, that is, which ends with an introduction.

The idea of various canonical forms is of course well-known in many mathematical contexts, but it has been objected to in this semantic context by some philosophers, for instance, by W. V. Quine. To extend the verificationistic idea to empirical areas outside mathematics, one must also explain the meaning of sentences in such areas in terms of proofs, or as we may prefer to say when we are outside of mathematics, in terms of verification; that is why we use the term 'verificationism' for this idea about how meaning is given. Again it is essential that we talk about canonical verifications and distinguish them from indirect verifications. The meaning of a statement is given by what counts as a canonical verification of it. We cannot exhaustively describe our indirect means of proving or verifying a statement, and therefore we cannot explain the meaning of a statement by saying what counts as a (direct or indirect) verification of it. If the meaning of a statement depended on all that counts as verification of it, then the meaning would change each time we found a new way of verifying the statement, which is contradicting a common experience. The distinction between canonical and indirect verifications is thus essential for a viable verificationism. However, Quine contests that we can legitimately make such a distinction; it would essentially amount to a distinction between the analytic and the synthetic. I shall not enter now into a discussion of this distinction, but I think it is a well-known phenomenon both in mathematical and non-mathematical practice that certain proofs or verifications are in need of no further justifications because they are clearly valid in virtue of what we mean by the terms involved—we cannot justify them except by referring to what we mean—while other proofs or verifications are such that when they are challenged we are willing to give further justifications (see Prawitz 1995).

3 Correctness of an assertion

That the meaning of a statement is determined by what counts as its canonical proof or verification is thus the essence of the verificationism that I have in mind. An indirect proof or verification can then be defined as something that shows that a direct verification can be given, could have been given, or will be possible to give; the tense to be chosen here depends on the tense of the sentence. In mathematics, where we do not need to take time into account, we can simply say that an indirect proof shows how a direct proof can be obtained, that is, it gives us a method that in principle allows us to find a canonical proof of the sentence in question.

Almost everyone agrees that in mathematics the conditions for the correctness of an assertion is that you know a (direct or indirect) proof of the sentence in question. It may of course happen that the sentence is true although your proof is wrong, or that you are just lucky in asserting a true sentence in spite of lacking all good reasons for thinking it to be true. We then normally say that although the sentence is true, its assertion was incorrect. Also a realist agrees to all that, so what has just been said is not something particular to verificationism. A realist does not give the same analysis of what a proof is as the verificationist, but regardless of how one analyses the notion of proof, one wants to say that to know a proof of a sentence is the (sufficient and necessary) condition for being right in asserting the sentence. It is true that (Jones 1998) claims in his chapter of this volume that knowledge of a proof is only a necessary condition for the correctness of an assertion. What he has in mind is, however, what we may call an alleged proof, while I have in mind proofs as just defined. If an alleged proof is found to be erroneous, then of course we have to withdraw an assertion based on that alleged proof—what I have called the condition for asserting the sentence was then not satisfied. The possibility of mistakes can never be ruled out. One may therefore say with Jones that to *think* that one has a proof is only a necessary condition for being right in asserting the sentence. But while we can never be absolutely sure that what we think is a proof is really a proof, it often happens, it seems, that we are in the possessions of real proofs and that accordingly we are right in asserting the corresponding sentence. I think that there need to be no disagreement on this point.

4 Truth

Even those who agree about the part of verificationism stated so far often disagree when it comes to the analysis of truth. One standpoint, which has sometimes been taken by Michael Dummett, is that a verificationist or intuitionist does not need a notion of the truth of a sentence different from that of the correctness of asserting a sentence. An assertion is on this view simply to be understood as a claim to the effect that a proof of the asserted sentence is known, thus a claim of knowing either a canonical proof or an indirect proof, which allows one to find a canonical one.

A realist certainly objects to that way of understanding a sentence. Although he may agree that one who asserts a sentence is obliged to know a proof of it, he maintains that this is not to be taken as the content of the sentence. He may express this by saying that you must make a distinction between what is guaranteed and what is said by an assertion. By asserting a sentence you guarantee that there is a proof of it, but that is not what the assertion says; the *content* of the sentence, what you say by asserting it, is simply that the sentence is true, not that you have a proof of it.

My standpoint is that the verificationist should say the same as the realist on this issue. It seems to be a misrepresentation of the assertion to think of its content as being that a proof has been found. It is to put too much in the content. Therefore, the verificationist also needs a notion of truth to be able to say that the content of the assertion is that the asserted sentence is true. As already indicated, Dummett is right in saying that the intuitionist is able to account from his point of view for the conditions for assertions of mathematical sentences in terms of just the notion of proof without invoking any notion of truth. My objection that we also need to account for what is said by an assertion, different from what is guaranteed by it, seems to have little weight if we confine ourselves to the practice of stating theorems within pure mathematics.

In applied mathematics the situation is different. Dummett agrees that the time element which is then added (as already hinted to in the above) has the effect that an indirect verification may only show that a direct verification could have been given; in other words, an indirect verification does not any more amount to the possession of a method for finding a direct verification. But I think that the problem is more general.

When mathematics is applied we typically have a sequence of inferences that involve both sentences belonging to pure mathematics and sentences of empirical content. The natural thing to say is then that such a sequence of inferences is correct if each inference preserves truth, where truth is a notion explained uniformly for all the sentences. Let me take a concrete example.

Up to quite recently I have been the owner of 52 sheep. From the premiss 'only 50 of my sheep are now in the meadow' made at that time one may thus correctly draw the conclusion 'two of my sheep are missing'. I can assure you that it is very tiresome to verify that there are 50 sheep in a field, because sheep constantly move around, and therefore you very easily lose track of them and have to start counting again. Given this, would it be correct to infer from the premiss 'only 50 of my sheep are now in the meadow' the conclusion 'there is now a very tired observer'? Clearly not. This conclusion seems, however, to follow if we take the premiss to mean the same as 'it has now been observed that only 50 of my sheep are now in the meadow'. If this has been observed there is indeed a very tired observer. This illustrates the strange consequences of understanding the content of the assertion to be that the asserted sentence has been verified; as already said, it is to put too much into the assertion. The unwarranted conclusion about the tired observer does not follow if we understand the premiss as only saying that the sentence in question is verifiable, that is, that it can be verified by a hypothetical suitably placed observer that only 50 of my sheep are now in the meadow. From the premiss so understood it still follows that two of my sheep are missing, that is, this is verifiable (it can be verified even by the same hypothetical observer).

My suggestion is that the content of the assertion should be analysed along the lines just suggested, that is, as being that the asserted sentence is verifiable. I also suggest that this is how truth is to be understood from a verificationist point of view, which then has the result that the content of an assertion is simply that the asserted sentence is true.

Also in pure mathematics there are phenomena that seem difficult to account for without such an objective notion of truth which does not refer to properties belonging to the speaker. We do not only assert sentences in mathematics, we also make conjectures and ask questions to ourselves. If we wonder whether there are infinitely many twin primes, we do not wonder whether this has been proved—we know already that it has not, that is why we wonder. We may of course wonder whether it will be proved, but a verificationist must be allowed to wonder not only that, but also whether it can be proved. Similarly, he may conjecture that there are infinitely many twin primes, and normally he is then not making the conjecture that it will be proved that there are infinitely many twin primes, which is a conjecture about future history. From a verificationist point of view the natural way to take the conjecture is to understand it as saying that it is provable that there are infinitely many primes. This may also be expressed by saying that there exists a proof of the proposition that there are infinitely many twin primes, where 'exists' is to be taken in a tenseless sense, not as implying that a proof has already been constructed by us.

I am thus arguing that even within a theory of meaning in terms of proofs (or verifications) we must make room for the possibility of entertaining ideas of provability or of abstract existence of proofs; it would be too narrow to construe our speech to be only about what is proved or about actual existence of proofs. Once we accept the notion of provability as legitimate, it is hardly controversial within verificationism that the truth of a proposition is to be identified with provability or existence of proofs. To avoid confusions it is here convenient to distinguish between actual and potential existence as Martin-Löf (1998) does, for example, in his chapter in this volume. We can then say that the correctness of an assertion requires the actual existence of a proof, while the truth of the asserted proposition requires only (and is identical with) the potential existence of a proof of the proposition. (As seen from his chapter in this volume, Martin-Löf is also analysing truth as the potential existence of a proof. However, he distinguishes between proofs and demonstrations, and has removed all epistemic content from the proofs. It is a demonstration that makes us know the existence of a proof or proof object. I say this only to indicate that there is a verbal agreement between us, but that our notions of proof are different; here I cannot enter into a deeper discussion of how they are related.)

But is it really permissible from a constructive point of view to speak of provability or (potential) existence of proofs? Dummett (1987; 1994) answers this question in the negative. He remarks that, if we identify truth with the existence of a proof and think of mathematical proofs as existing independently of our hitting upon them, then it is hard to see how we can resist the idea that a proof of a statement either exists or fails to exist, and since the non-existence of a proof must be identified with the falsity of the statement in question, we are then back to the law of bivalence and full realism.

Although the idea of proofs existing independently of our hitting upon them certainly contains a flavour of realism, I do not think that it amounts to a full step to realism. I want to give two reasons for thinking so. Firstly, proofs as here understood are something that in principle can be known by us, and hence there is no talk about in principle unknowable proofs. Secondly, I do not see why the disjunction 'either there exists a proof of A or there does not exist a proof of A' must be taken in a classical way. Although we think of the proofs as having some kind of existence even before we find them, an intuitionist may still maintain that to assert the disjunction that either there is or there is not a proof of A requires that we know how to find a verification either of the existence of a proof of A or of the non-existence of a proof of A. For an arbitrary A we do not know how to find such a verification, and we should then have no difficulty in resisting the thought that the disjunction in question is true.

5 Objectivity

Most of us, including myself, are convinced that mathematics is objective. Something does not become correct in mathematics because we hold it to be correct. It is conceivable that we all think that a theorem has been proved, but as a matter of fact the proof is erroneous and the asserted proposition is false. A realist picture of some domain such as mathematics is commonly held to enforce this kind of objectivity. According to the realist picture a sentence is understood as saying something about an independently existing world, and it is true if the world is as the sentence says it is. Whether a sentence is true is thus something thoroughly objective, which has nothing to do with us but depends only on how it is in the world. The problem that I want to take up is how objectivity comes out when one takes a verificationist point of view. Can objectivity then be still maintained?

Martin-Löf (1987) answers this question positively emphasizing the objectivity of the notion of the validity of a proof, and contrasting his position to that of a subjectivist who lacks such a notion and who can only say that a judgement is evident for him—in case of conflicts where an opposite judgement is evident for someone else, the subjectivist has nothing to appeal to that allows him to think of it as a conflict that should be resolved.

Similarly, in discussions of Wittgenstein's philosophy, Dummett (1978; 1987) rejects the subjectivist view which he ascribes to Wittgenstein that there is nothing more to the validity of a proof than our treating it as a proof. On such a view the practice of deduction loses its point, Dummett says, because the point of a deduction is not just to settle issues in one way or the other but to prove propositions that are true in virtue of the meaning that we have already conferred upon the terms involved. Accordingly, Dummett rejects not only the platonistic picture that the mathematician discovers how it is in the world of mathematical objects, but also the subjectivist picture that we freely create that world. Between these two metaphysical pictures, Dummett interposes a third intermediate picture

> ... of objects springing into being in response to our probing. We do not make the objects but must accept them as we find them ...; but they were not already there for our statements to be true or false of before we carried out the investigation which brought them into being (Dummett 1978, p. 185).

I agree with the view that a philosophy of mathematics that cannot account for the objectivity of mathematics is amiss. But is it so clear that the objectivity of mathematics can be maintained when one takes a verificationist point of view? Of course, we can say from that point of view that truths are objective because a sentence is true in virtue of the fact that there is a proof of it. But what is it that makes proofs objective?

As defined here, something is a proof in mathematics if it is either a canonical proof or a method for finding a canonical proof. We may claim that once we have laid down what counts as canonical proofs, it is a factual matter whether an alleged proof amounts to such a canonical proof. If it is not a canonical proof, then it is again a factual matter whether the alleged proof yields a method for finding a canonical proof. Hence it should be clear that it is not our treating it as a proof that makes it a proof. This seems to be a reasonable claim. It makes something a proof in virtue of the meaning of the expressions involved, which is also reasonable. But it also seems to imply that the question of whether something is a proof is fixed when the meanings are given, that is, when it is given what counts as a canonical proof. From this it is natural to conclude that already, before a proof of a sentence is found, it is determined that there is such a proof. Provability, which I want to identify with truth, becomes in this way something objective.

My point is that the same features which make proof and to be proved objective notions also make provability objective. If we discard the notion of provability, maintaining that before a sentence is proved it is not determined whether it is provable, then it seems that we pull away the grounds for the objectivity of proofs. I am thus concluding that, if we are to reject subjectivity, then we should maintain that the question of whether something is a proof of a given sentence is objectively determined by features which also determine whether the sentence is provable and which determine this already before it is proved. Having thus accepted the notion of provability, we should obviously identify the truth of a sentence not with it being proved but with it being provable.

Bibliography

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

Dales, H. G. (1998). The mathematician as a formalist. *This volume*, 181–200.

Dummett, M. A. E. (1978). The philosophical significance of Gödel's theorem. Reprinted in his *Truth and other enigmas*, pp. 186–201. Harvard University Press, Cambridge, Massachusetts.

Dummett, M. A. E. (1987). Reply to Dag Prawitz. In *Michael Dummett: contributions to philosophy* (ed. B. Taylor), pp. 281–6. Martinus Nijhoff, Dordrecht.

Dummett, M. A. E. (1993*a*). Wittgenstein on necessity: some reflections. In *The seas of language*, pp. 446–61. Clarendon Press, Oxford.

Dummett, M. A. E. (1994). Reply to Prawitz. In *The philosophy of Michael Dummett* (ed. B. McGuinness and G. Oliveri), pp. 292–8. Synthèse Library, Vol. 239. Kluwer Academic, Dordrecht.

Gentzen, G. (1934–35). Untersuchungen über das logisches Schliessen. *Mathematische Zeitschrift*, **39**, 176–210, 405–31.

Gentzen, G. (1938). Neue Fassung des Wiederspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der Exakten Wissenschaften*, **4**, 19–44.

Jones, V. F. R. (1998). A credo of sorts. *This volume*, 203–214.

Martin-Löf, P. (1971). Hauptzatz for the intuitionistic theory of iterated inductive definitions. In *Proceedings of the 2nd Scandinavian Logic Symposium* (ed. J.E. Fenstad), pp. 179–216. North-Holland, Amsterdam.

Martin-Löf, P. (1974). *Intuitionistic type theory.* Bibliopolis, Napoli.

Martin-Löf, P. (1987). Truth of a proposition, evidence of a judgement, validity of a proof. *Synthèse*, **73**, 407–20.

Martin-Löf, P. (1998). Truth and knowability: on the principles C and K of Michael Dummett. *This volume,* 105–14.

Prawitz, D. (1965). *Natural deduction: a proof-theoretical study.* Almqvist & Wiksell, Stockholm.

Prawitz, D. (1971). Ideas and results in general proof theory. In *Proceedings of the 2nd Scandinavian Logic Symposium* (ed. J.E. Fenstad), pp. 235–307. North-Holland, Amsterdam.

Prawitz, D. (1995). Quine and verificationism. *Inquiry*, **37**, 487–94.

Department of Philosophy
Stockholm University
S-106 91 Stockholm
Sweden
email: Dag.Prawitz@philosophy.su.se

3

Constructive truth in practice

Douglas S. Bridges

1 What is constructive mathematics?

In this chapter, which has evolved over the last ten years to what I hope will be its perfect Platonic form, I shall first discuss those features of constructive mathematics that distinguish it from its traditional, or *classical*, counterpart, and then illustrate the practice of that distinction in aspects of complex analysis whose classical treatment ought to be familiar to a beginning graduate student of pure mathematics.

My experience shows that a typical mathematician believes that constructive mathematics is characterized by either a rejection of the law of excluded middle from logic or else a rejection of the full axiom of choice from set theory. In fact, although some authors (notably Richman (1996)) seem to endorse the former characterization, the pioneers of constructivism—Brouwer, Markov, Bishop—all arrived at their rejection of the law of excluded middle (and hence, implicitly, at a rejection of the axiom of choice—see later in this chapter) as a consequence of their insistence that the phrase *there exists* be interpreted strictly as *we can construct.* Thus the key feature of constructive mathematics is the identification

EXISTENCE ≡ COMPUTABILITY

At this point one might ask what is wrong with the classical computability theory based on recursive functions: does not that theory provide a suitably constrained framework for the discussion of questions of constructivity, outside which we can continue to handle 'idealistic' existence theorems with impunity?

Consider the following example of a function f from the set $\mathbb{N}$ of natural numbers to itself:

$$f(n) = \begin{cases} 0 & \text{if the Goldbach conjecture is false,} \\ 1 & \text{if the Goldbach conjecture is true.} \end{cases}$$

(The Goldbach Conjecture states that every even integer greater than 2 is a sum of two primes.) In classical mathematics this is regarded as a computable function, since there exists (in the idealistic sense—it is absurd that there not exist) an algorithm that, applied to any natural number n, outputs $f(n)$. In fact, there are two algorithms—one, $\mathcal{A}_0$, that always outputs 0, and one, $\mathcal{A}_1$, that always outputs 1—one of which must, classically, compute f.

Now, I find it strange to describe a function $f : \mathbb{N} \to \mathbb{N}$ as computable when, with our present state of mathematical knowledge, we cannot even compute its value at the input 0. The following example is, however, even stranger. Define $g : \mathbb{N} \to \mathbb{N}$ by

$$g(n) = \begin{cases} 0 & \text{if the Continuum Hypothesis is false,} \\ 1 & \text{if the Continuum Hypothesis is true.} \end{cases}$$

(The Continuum Hypothesis says that $2^{\aleph_0} = \aleph_1$.) A platonist would certainly have no trouble accepting this as a good definition of a function, and would have to say that g, like f, is computable; but this time we have no possibility whatsoever of determining the value $g(0)$ within ZFC (Zermelo–Fræenkel set theory plus the axiom of choice), the normal setting for classical mathematics.

Is it, then, sensible to call such functions as f and g 'computable'? The constructivist would say

> *No. It only makes sense to call f computable if we can decide which of the two algorithms $\mathcal{A}_0$ and $\mathcal{A}_1$ computes f; and it only makes sense to call g computable if we can decide which of $\mathcal{A}_0$ and $\mathcal{A}_1$ computes g. In other words, we are only justified in calling f computable if we can decide the Goldbach conjecture; and we are only justified in calling g computable if we can decide the Continuum Hypothesis—which, of course, we will never be able to do unless we step outside* ZFC.

Unfortunately, classical logic is not refined enough to enable us to distinguish between computability in the constructivist's stronger sense—we can pinpoint the algorithm that computes the function—and the notion of computability that includes our freakish, continuum hypothesis-based, function g. Fortunately, there is a logic that makes it easier for us to avoid bringing such examples into our mathematics: I refer, of course, to *intuitionistic logic*, abstracted by Heyting from the practice of Brouwer's intuitionistic mathematics.

For example, in that logic:

- $P \vee Q$ holds if and only if we have either a proof of P or a proof of Q;
- $\exists x\, P(x)$ holds if and only if we have an algorithm for constructing x, *and* one for verifying that $P(x)$ holds;
- as we shall see later, even for a decidable property $P(n)$ of natural numbers n,

  $$\forall n\, P(n) \vee \neg\forall n\, P(n)$$

  need not hold; and so
- the *law of excluded middle*, $P \vee \neg P$, does not hold.

There are grounds for arguing that intuitionistic logic is more natural than classical logic in the study of computability.[1] Certainly, as Richman (1983) has shown, fundamental results in elementary recursion theory, such as the *s-m-n* Theorem, Rice's Theorem, and the Recursion Theorem, can be proved using intuitionistic logic. However, the restriction to that logic would lose us many important results, the full form of the Speed-up Theorem being a case in point (see (Bridges 1994*a*, (6.20) and (6.26))).

2 Varieties of constructive mathematics

Meaningful distinctions deserve to be maintained. (Bishop 1973)

There are three major varieties of constructive mathematics that are currently the domain of serious research activity. In making this claim I do not wish to belittle some of the other approaches to constructivity, such as the ultra-intuitionistic, finitist programme of Yessenin-Volpin (1970); but such approaches are, at present, of marginal significance.

The first variety, and historically the oldest, is Brouwer's intuitionistic mathematics (INT). For several decades after the publication of his thesis (Brouwer 1907), Brouwer developed mathematics based on his philosophy of *intuitionism*, a philosophy that led him to a number of concepts and principles which, on first reading, appear to contradict classical mathematics (CLASS). For example, a theorem of INT states that

> (*) *Every function from* $[0,1]$ *to the real line* $\mathbb{R}$ *is uniformly continuous.*

The apparent absurdity of this statement is, however, illusory, as is suggested by the following more careful re-statement of it:

> (**) *Every intuitionistically defined function from the intuitionistic interval* $[0,1]$ *to the intuitionistic real line is, intuitionistically, uniformly continuous.*

In fact, there is a strong case for saying that, except at certain levels of formalism, INT and CLASS are so divergent that it is not possible to capture fully the spirit and meaning of intuitionistic statements, such as (**), within a classical framework.

For more details about INT, see (Dummett 1977), (van Dalen 1981), (Troelstra and van Dalen 1988), and (Bridges and Richman 1987).

The second variety is the recursive constructive mathematics (RUSS) of the Russian school founded by Markov. In this variety, the fundamental objects are natural numbers and recursive functions, and the logic is intuitionistic. More complex objects are represented by Gödel numbers; so, for example, a function between sets of recursive reals is represented by a function between the corresponding sets of Gödel numbers. This leads us to a theorem that *prima facie* contradicts the Uniform Continuity Theorem (Dieudonné 1960, (3.16.5)) of classical mathematics:

> ($\natural$) *There exists a pointwise continuous function* $f : [0,1] \to \mathbb{R}$ *that is not uniformly continuous.*

Once again, the apparent contradiction disappears when we interpret our statement more carefully:

($\natural\natural$) *There exists a recursive function, mapping the recursive interval* $[0, 1]$ *into the recursive real line, that is recursively pointwise continuous but not recursively uniformly continuous.*

Since the recursive reals form a classically countable set, ($\natural\natural$) is perhaps not surprising; yet it is hard to prove within RUSS, in which the recursive real line is not *recursively* countable (effectively enumerable).

My third variety is the constructive mathematics (BISH) first described in the late Errett Bishop's ground-breaking monograph (Bishop 1967); see also (Bishop and Bridges 1985). Without committing himself either to the philosophical/metaphysical principles espoused by Brouwer or to the constrictive framework of recursive function theory, and treating *algorithm*, or *finite routine,* as a primitive, undefined notion, Bishop was able to confound Hilbert's prediction that constructive mathematics would be unable to produce deep results: single-handedly, he obtained constructive analogues of such cornerstones of analysis as the Hahn–Banach theorem, the spectral theory of self-adjoint operators, abstract measure theory (including Haar measure on locally compact groups), and the elements of Banach algebra theory. I will try to give the flavour of Bishop's mathematics with examples from complex analysis in the second half of this chapter.

There is a strong case for regarding BISH as the constructive core of mathematics, since every theorem of BISH is also a theorem of INT, RUSS, and CLASS; at a certain level of formalism, BISH is consistent with each of those three varieties of mathematics, which we can therefore regard as models of BISH.

This observation leads to some interesting independence results. For example, each of the statements (*) and ($\natural$) mentioned above is independent of BISH. For if (*) were provable in BISH, then it would be a theorem of RUSS, which would contradict ($\natural$); whereas if ($\natural$) were provable in BISH, then it would be a theorem of INT, which would contradict (*). Similarly, each of the statements

Every function from $[0, 1]$ *to* $\mathbb{R}$ *is Lebesgue integrable*

(a theorem of INT) and

There exists a bounded, pointwise continuous function from $[0, 1]$ *to* $\mathbb{R}$ *that is not Lebesgue measurable*

(a theorem of RUSS) is independent of BISH. To prove the first of these statements in BISH, we would need to add some principles like those invoked by Brouwer; to prove the second, we would need to add to BISH something like Church's thesis (Bridges and Demuth 1991).

There is another view of constructive mathematics that is worth mentioning here. Fred Richman has come to regard constructive mathematics as mathematics that deals with the normal objects of CLASS, but rejects the law of excluded middle; see (Richman 1996). I have some sympathy with this view, but would hold that the rejection of the law of excluded middle derives from the requirement that all methods employed in constructive mathematics be algorithmic; in other words, *algorithmic method drives our choice of logic and techniques.*

3 Omniscience principles

From now on, I shall only consider constructively defined objects. So when I speak of a 'binary sequence', I shall assume without comment that this sequence is presented constructively; in other words, there is a finite routine which generates the terms of the sequence one by one.

One feature of constructive mathematics, originating with Brouwer, that often irritates the classical mathematician is the use of what Bishop called *omniscience principles* to demonstrate that certain classical results are essentially non-constructive. Among these omniscience principles, which are really weak forms of the law of excluded middle, are the following:

> *The limited principle of omniscience* (LPO): If (a_n) is a binary sequence, then either $a_n = 0$ for all n or else there exists n with $a_n = 1$.
>
> *The lesser limited principle of omniscience* (LLPO): If (a_n) is a binary sequence with at most one term equal to 1, then either $a_{2n} = 0$ for all n or else $a_{2n+1} = 0$ for all n.

Both of these principles are false if interpreted recursively, *even with classical logic* (see (Bridges 1994*a*, Chapter 4)). In other words, their non-constructive nature stems not from the underlying logic but from their failure in the recursive model.

The following omniscience principle is not a consequence of the law of excluded middle, and holds in the classical recursive model:

> *Markov's principle* (MP): If (a_n) is a binary sequence for which it is contradictory that all terms be zero, then there exists n such that $a_n = 1$.

This principle

- represents an unbounded search;
- is accepted in RUSS;
- is contradicted by Brouwer's theory of the creating subject (Heyting 1971);
- is not used in BISH.

Although many classical theorems are equivalent, constructively, to Markov's principle, applications of that principle can often be avoided by careful use of the completeness of an appropriately chosen metric space. Such applications of completeness often also enable us to prove constructively propositions that are trivial consequences of LPO or LLPO. I shall return to this shortly.

By a *Brouwerian counterexample* to a classical proposition P we mean a constructive proof that P entails some omniscience principle; if the omniscience principle is MP, we often speak of a *Markovian counterexample*, rather than a Brouwerian one.

Brouwerian counterexamples are not counterexamples in the usual sense, but provide *strong evidence* that certain classical propositions will never be proved constructively. (Note, however, that when the omniscience principle involved in a Brouwerian counterexample to a proposition P is one, such as LPO or LLPO, that is false in the recursive model, then that Brouwerian counterexample can be turned into a proof that the recursive interpretation of P is false.) As we shall see when I talk about the Riemann Mapping Theorem, a Brouwerian counterexample to P can also provide positive information, by suggesting additional constructive hypotheses that will lead to a constructive counterpart to P.

Here are two Brouwerian counterexamples that pertain to the constructive real number line $\mathbb{R}$.

(1) $\forall x \in \mathbb{R}$ $(x = 0$ *or* $x \neq 0)$ *implies* LPO.

Proof: First note that $x \neq 0$ means that we can compute a rational number between 0 and x. Let (a_n) be a binary sequence, and let x be the real number whose binary expansion is $0 \cdot a_1a_2a_3\cdots$. This is certainly a well defined constructive real number, as we can specify it to any degree of accuracy simply by computing enough terms of the binary sequence. If $x = 0$, then $a_n = 0$ for each n. If $x \neq 0$, then we can compute a positive integer k such that $x > 2^{-k}$; by testing the terms $a_1, \ldots, a_{k+1}$, we can then find one that equals 1. □

(2) $\forall x \in \mathbb{R}$ $(x \geq 0$ *or* $x \leq 0)$ *implies* LLPO.

Proof: In this case, given a binary sequence (a_n) with at most one term equal to 1, we need only consider the real number

$$x \equiv \frac{a_1}{2} - \frac{a_2}{4} + \frac{a_3}{8} - \frac{a_4}{16} + \cdots .$$

If $a_N = 1$ and N is even, then

$$x = -a_N/2^N < 0\,;$$

so if $x \geq 0$, then $a_n = 0$ for all even n. Similarly, if $x \leq 0$, then $a_n = 0$ for all odd n. □

Now, after these two Brouwerian examples you may be thinking that it must be impossible to do much with the constructive real line $\mathbb{R}$. Fortunately, such thoughts are wrong: there are constructive principles that enable us to get round the inadmissibility of

$$\forall x \in \mathbb{R}\, (x = 0 \ \text{ or } \ x \neq 0)$$

and

$$\forall x \in \mathbb{R}\, (x \geq 0 \ \text{ or } \ x \leq 0)$$

within BISH. Two commonly used constructive principles of this type are

If $a > b$, then, for each $x \in \mathbb{R}$, either $a > x$ or $x > b$

and

If $x > 0$ is contradictory, then $x \leq 0$.

The first of these is proved by taking sufficiently close rational approximations to a, b, and x. The second follows from the formal constructive definition of *real number* in Bishop (1967, Chapter 2). (Note, incidentally, that the statement

$$\forall x \in \mathbb{R}\ (\neg(x \geq 0) \Rightarrow x < 0)$$

is equivalent to Markov's Principle, and is therefore not used in BISH.)

There is one further matter that I would like to tidy up here: the status of the axiom of choice within BISH. The following result was proved in (Goodman and Myhill 1978), but was surely known to Bishop when he wrote his original monograph (see Bishop (1967, page 58, Problem 2)):

The axiom of choice implies the law of excluded middle.

To show this, let P be any constructively meaningful statement, and define the set A to consist of the two elements 0 and 1, together with the equality relation

$$0 = 1 \text{ if and only if } P \text{ holds}.$$

(We could have defined A in more classical terms as a set of equivalence classes under the equivalence relation

$$0 \sim 1 \text{ if and only if } P \text{ holds},$$

but it is more in keeping with Bishop's approach to proceed as we have done.) Let B be the set $\{0, 1\}$ with the standard equality, and let

$$S \equiv \{(0,0),\ (1,1)\} \subset A \times B,$$

where the equality on S is derived in the usual way from those on A and B:

$$(x, y) = (x', y') \text{ if and only if } \ x = x' \text{ in } A \text{ and } \ y = y' \text{ in } B.$$

Suppose that there exists a function $f : A \to B$ such that $(x, f(x)) \in S$ for all $x \in A$. If $f(0) = 1$ or $f(1) = 0$, then $0 = 1$, and hence P holds; if $f(0) = 0$ and $f(1) = 1$, then $\neg(0 = 1)$, and hence P is false. Thus we have derived $P \vee \neg P$.

I should point out here that most constructive mathematicians freely use the weaker axioms of countable choice and dependent choice.

4 Completeness in constructive analysis

Another foundation stone of classical analysis that crumbles to dust in the constructive framework is the expression of the completeness of $\mathbb{R}$ in the *least upper bound principle*: every non-empty subset of $\mathbb{R}$ that is bounded above has a least upper bound. Indeed, we cannot even guarantee the construction of the least upper bound of an increasing binary sequence. To see this, let (a_n) be any increasing binary sequence, and suppose that $s \equiv \sup a_n$ exists. Then either $s > 0$, in which case there exists n with $a_n = 1$, or else $s < 1$; in the latter case it is impossible for any a_n to equal 1, so $a_n = 0$ for all n. It now easily follows that the classical least upper bound principle implies LPO.

There is, however, an extremely useful constructive version of the least upper bound principle—namely,

> $\mathrm{LUB_c}$: *Let S be a subset of $\mathbb{R}$ that is non-empty and bounded above. In order that S have a supremum, it is necessary and sufficient that for all real numbers α, β with $\alpha < \beta$, either there exists $s \in S$ such that $s > \alpha$ or else β is an upper bound of S.*

(*Non-empty* here means that we can construct an element of S; in other words, S is *inhabited.*) This can be proved using Bishop's definition of real numbers as special Cauchy sequences of rationals, together with the 'Cauchy sequence completeness' of the constructive real line; see (Bridges 1994*b*). It can also be used as the axiom of completeness in an axiomatic development of the constructive real line (Bridges 1998).

To illustrate the use of $\mathrm{LUB_c}$, we prove that

> *A totally bounded subset of* $\mathbb{R}$ *has a supremum.*

Let $S \subset \mathbb{R}$ be totally bounded; so for each $\varepsilon > 0$ there exists a finite ε-approximation to S—that is, a finite set $F \subset S$ such that for each $s \in S$ there exists $x \in F$ with $|s - x| < \varepsilon$. Let α, β be real numbers with $\alpha < \beta$, and set $\varepsilon \equiv (\beta - \alpha)/3$. Let $\{x_1, \dots, x_n\}$ be a finite ε-approximation to S, and choose N such that

$$x_N > \sup\{x_1, \dots, x_n\} - \varepsilon .$$

(This supremum exists since it applies to a finite set.) Either $\alpha < x_N$ or $x_N < \alpha + \varepsilon$. In the latter case, if $s \in S$ and $|s - x_k| < \varepsilon$, then

$$s \le x_k + \varepsilon < x_N + 2\varepsilon < \alpha + 3\varepsilon = \beta ,$$

so β is an upper bound of S. Thus $\sup S$ exists, by $\mathrm{LUB_c}$.

I have already commented that Brouwerian counterexamples have the positive role of suggesting hypotheses—in many cases, ones which hold trivially in CLASS—which, when added to the normal ones of a classical, but essentially non-constructive, theorem, enable the augmented theorem to be proved constructively. For example, let a be a real number and define a uniformly continuous function $f : [0,1] \to \mathbb{R}$ such that

- $f(0) = -1,\ f(1) = 1,\ f(1/3) = -a,\ f(2/3) = a$;
- f is linear in each of the intervals $\left[0, \frac{1}{3}\right], \left[\frac{1}{3}, \frac{2}{3}\right], \left[\frac{2}{3}, 1\right]$.

It is not hard to show that, if there exists $x \in [0,1]$ such that $f(x) = 0$, then either $a \ge 0$ or $a \le 0$. It follows that

> *The classical intermediate value theorem implies* LLPO.

Now, a little reflection on this function f should convince you that if we could rule out the possibility that its graph ever flattens, then we could construct a zero of f. We can accomplish this, using an approximate interval-halving argument and therefore, implicitly, the completeness of $\mathbb{R}$, if we add the hypothesis that

f is *locally non-zero*: that is, for each $x \in (0,1)$ and each $r > 0$ there exists x' in the interval $(x - r, x + r)$ such that $f(x') \neq 0$. We then obtain the following proposition, which is one of several constructive substitutes for the classical intermediate value theorem:

> *If f is uniformly continuous and locally non-zero on the unit interval $[0,1]$, and $f(0)f(1) < 0$, then there exists x with $f(x) = 0$.*

Here is another substitute, in which we retain the weaker classical hypotheses but also weaken the constructive conclusion:

> *If f is uniformly continuous on $[0,1]$, $f(0)f(1) < 0$, and $\varepsilon > 0$, then there exists x with $|f(x)| < \varepsilon$.*

For more on the Intermediate Value Theorem, see (Bishop and Bridges 1985, Chapter 2) and (Bridges and Richman 1987, Chapter 3).

This is an example of the *bifurcation* of a classical theorem, a phenomenon in which the theorem has several constructively inequivalent constructive counterparts, each of which is classically equivalent to the others. We shall meet another example of this shortly, when we look at Picard's Theorem.

I would now like to return to a matter alluded to earlier: namely, how we can use completeness to avoid appealing to omniscience principles. Consider the linear space

$$\mathbb{R}a \equiv \{ax : x \in \mathbb{R}\} ,$$

where a is a given real number. It is trivial that

- if $a = 0$, then $\mathbb{R}a = \{0\}$ and is 0-dimensional;
- if $a \neq 0$, then $\mathbb{R}a = \mathbb{R}$ and is 1-dimensional;
- if $\mathbb{R}a$ is finite-dimensional, then either $a = 0$ or $a \neq 0$, and $\mathbb{R}a$ is closed in $\mathbb{R}$.

In particular, it follows from the third of these statements and our first Brouwerian counterexample that, if $\mathbb{R}a$ is finite-dimensional for any $a \in \mathbb{R}$, then LPO holds; so we have no hope of proving that $\mathbb{R}a$ is finite-dimensional in general. However, we can prove that

> *If $\mathbb{R}a$ is closed in $\mathbb{R}$, then it is finite-dimensional.*

To this end, assuming that $\mathbb{R}a$ is closed in $\mathbb{R}$, we construct a decreasing binary sequence (λ_n) such that

$$\begin{aligned} \lambda_n = 1 \quad &\Rightarrow \quad |a| < 1/n^2 , \\ \lambda_n = 0 \quad &\Rightarrow \quad |a| > 1/(n+1)^2 ; \end{aligned}$$

this construction is possible since either $1/n^2 > |a|$ or $|a| > 1/(n+1)^2$. The completeness of $\mathbb{R}$ ensures that the series $\sum_{n=1}^{\infty} \lambda_n a$ converges by comparison with $\sum_{n=1}^{\infty} 1/n^2$. Since $\mathbb{R}a$ is a closed subset of $\mathbb{R}$, the sum of the series $\sum_{n=1}^{\infty} \lambda_n a$ has the form ξa for some $\xi \in \mathbb{R}$. Choose $N > |\xi|$, and consider λ_N. If $\lambda_N = 0$, then $a \neq 0$ and we are finished; so we may assume that $\lambda_N = 1$. Suppose there exists $m \geq N$ such that $\lambda_{m+1} = 1 - \lambda_m$. Then $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 1$, and $\lambda_k = 0$ for all $k > m$; so

$$\xi a = \sum_{n=1}^{\infty} \lambda_n a = \sum_{n=1}^{m} \lambda_n a = ma \,,$$

and therefore $\xi = m$. (Note that, as $\lambda_{m+1} = 0$, $|a| > 1/(m+2)^2$ and we can divide by a.) This is absurd, as $m \geq N > |\xi|$. We conclude that if $\lambda_N = 1$ for our special choice of N, then $\lambda_m = 1$ for all $m \geq N$ and hence for all m; it follows that $|a| < 1/m^2$ for all m, and therefore that $a = 0$. Thus, by a careful construction of a series whose convergence depends on the completeness of $\mathbb{R}$, and an equally careful estimation using its sum, we have been able to show that if $\mathbb{R}a$ is closed, then either $a = 0$ or $a \neq 0$.

Now, this may look like a rather uninteresting example, since spaces of the type $\mathbb{R}a$ do not occur very often in advanced analysis. But our result about $\mathbb{R}a$ suggested, and is a special case of, the more general theorem,

> *A Banach space with a compact generating set is finite-dimensional* (Richman *et al.* 1982),

whose proof requires several applications of completeness similar to, and in one case generalizing, the one we have just used. In turn, this theorem enables us to prove that if the range of a compact linear mapping between normed spaces is complete, then that range is finite-dimensional—a result whose standard classical proof depends on a version of the Open Mapping Theorem that is not known to be constructive (see Theorem 4.18 of (Rudin 1973), and (Bridges *et al.* 1989)).

Those of us working in constructive analysis have found many situations where an application of completeness similar to the one used above has enabled us to circumvent omniscience principles. The completeness usually has to be added to the hypotheses of what would otherwise be a trivial classical theorem. In our discussion of $\mathbb{R}a$, although the completeness of $\mathbb{R}$ is used to establish the convergence of the series $\sum_{n=1}^{\infty} \lambda_n a$, we really need $\mathbb{R}a$ to be complete in order to ensure that the series converges to a sum that belongs to $\mathbb{R}a$; the required completeness is implicitly contained in the hypothesis that $\mathbb{R}a$ is a closed subset of $\mathbb{R}$.

Among many interesting constructive theorems whose proofs use such applications of completeness are the following:

- *If f is a non-negative Lebesgue integrable function that is positive throughout a set of positive measure, then* $\int f > 0$ (Bishop and Bridges 1985, Chapter 6, (4.13)).
- *A linear mapping T of a normed space X onto a Banach space is* well-behaved, *in the sense that, if $x \in X$ is distinct from each element of the kernel of T, then $Tx \neq 0$* (Bridges and Ishihara 1990).
- *Let F be a finite-dimensional subspace of a normed space X, and let $a \in X$*

have at most one best approximation in F, *in the following sense: if* x, x' *are distinct elements of* F, *then*

$$\max\{\|a - x\|, \|a - x'\|\} > \rho(a, F) \equiv \inf\{\|a - y\| : y \in F\} .$$

Then there exists a unique element $b \in F$ *such that* $\|a - b\| = \rho(a, F)$ (Bridges 1981).

5 Complex analysis

I would now like to turn from the more general, and occasionally more negative, features of constructive mathematics that we have focused on so far, to demonstrate its positive features by means of examples drawn from complex analysis. All our subsequent discussion will be set in the context of BISH.

First, let us look at the classical *Jordan Curve Theorem*:

> *If* J *is a Jordan curve in* $\mathbb{C}$, *then* $\mathbb{C}\backslash J$ *is the union of two disjoint open connected sets.*

(A *Jordan curve* is a one-one, uniformly continuous mapping of the unit circle into $\mathbb{C}$, with a uniformly continuous inverse.) For various reasons, the standard classical proofs of this theorem fall down constructively. However, Berg *et al.* (1975) have produced the following constructive Jordan Curve Theorem:

> *For any two points* a, b *bounded away from a Jordan curve* J, *either* a *and* b *can be joined by a polygonal path bounded away from* J, *or else the winding numbers of* J *with respect to* a *and* b *differ by* 1.

In other words, we have an algorithm which, when provided with

the data specifying a Jordan curve J,
a pair of complex numbers a, b, and
a proof that a and b are bounded away from J,

will

either construct a polygonal path $p : [0, 1] \to \mathbb{C}$ joining a and b, compute the number

$$\rho \equiv \inf\{|x - p(t)| : x \in J, 0 \leq t \leq 1\} ,$$

and show that $\rho > 0$,

or compute the winding numbers of J with respect to a and b, and show that these are unequal.

In the first case, a and b both belong to the same component (*inside* or *outside*) of $\mathbb{C}\backslash J$; in the second, they are in different components.

There remains, however, the following significant constructive question: starting at any point on the curve, can we tell in which direction we should move to get inside J? Could it be the case that, if J is sufficiently convoluted (without crossing itself), we simply cannot tell how to get inside it at certain points? Again, Berg *et al.* (1975) have given us the answer:

> *If $z \in J$, then there exist z_0 and z_1, bounded away from J but arbitrarily close to z, such that the winding numbers of J with respect to z_0 and z_1 are different.*

Moreover, comparison of these winding numbers with that of a point far removed from, and therefore clearly outside, J will enable us to tell which of z_0 and z_1 is inside, and which is outside, our curve.

For my next example, let f be a differentiable complex function on the punctured disc

$$U \equiv \{z \in \mathbb{C} : 0 < |z - z_0| < r\}.$$

Picard's Theorem (sometimes called the *Great Picard Theorem*) states that

> *If f has an essential singularity at z_0, then the range of f is either $\mathbb{C}$ or else $\mathbb{C}$ with one point omitted.*

This result was published by Emile Picard in 1879. One hundred years later, I and my co-workers at New Mexico State University carried out a constructive analysis (Bridges *et al.* 1982) in which we showed, by means of a Brouwerian counterexample, that the classical form of the theorem is essentially nonconstructive. We then proved the following constructive version of the theorem:

> *If f has an essential singularity at z_0, and if ζ, ζ' are distinct complex numbers, then there exists z in U such that either $f(z) = \zeta$ or else $f(z) = \zeta'$.*

A simple application of the law of excluded middle shows that this version is classically equivalent to Picard's original theorem.

Now, there is another classically equivalent version that turns out also to be provable within BISH:

> *If the range of f omits two complex values, then f has a pole of determinate order at z_0.*

However, our two constructive versions are definitely not equivalent within BISH, as they embody totally different algorithms. In the first case, we have an algorithm which, when applied to data consisting of

the Laurent expansion $\sum_{n=-\infty}^{\infty} a_n(z - z_0)^n$ of f about z_0,

a strictly increasing sequence of positive integers $(n_k)_{k=1}^{\infty}$ such that $a_{-n_k} \neq 0$ for each k, and

two distinct complex numbers ζ, ζ',

computes a point z of U and either shows that $f(z) = \zeta$ or else shows that $f(z) = \zeta'$. So the first algorithm enables us to solve equations of the type

$$f(z) = \zeta\,.$$

On the other hand, the second constructive Picard Theorem produces an algorithm which, when applied to data consisting of

the Laurent expansion $\sum_{n=-\infty}^{\infty} a_n(z - z_0)^n$ of f about z_0,
two distinct complex numbers ζ, ζ',
a proof that $f(z) \neq \zeta$ for all $z \in U$, and
a proof that $f(z) \neq \zeta'$ for all $z \in U$,

computes a certain integer ν, shows that $a_\nu \neq 0$, and shows that $a_n = 0$ for all $n < \nu$. So our second algorithm computes the least index of a non-zero term of the Laurent expansion of f.

My last example from complex analysis is the famous *Riemann Mapping Theorem*:

> *If U is a proper, open, simply connected subset of $\mathbb{C}$, then there exists an analytic equivalence of U with the open unit disc D: that is, a one-one analytic mapping f* (a Riemann mapping) *of D onto U with analytic inverse* (Rudin 1970, Theorem 14.8).

This time, in order to provide motivation for the hypotheses of the constructive Riemann Mapping Theorem, I shall give the details of a Brouwerian counterexample to the classical theorem. Given a binary sequence (a_n), define $U \equiv \bigcup_{n=1}^{\infty} S_n$, where

$S_n = D$ if $a_n = 0$, and
S_n is the open disc with centre 0 and radius 2 if $a_n = 1$.

Then U is open and simply connected, and is clearly a proper subset of $\mathbb{C}$. Suppose that f is an analytic equivalence of U with D; we may assume that $f(0) = 0$. Now, either $|f'(0)| > 1$ or $|f'(0)| < 2$. (Recall that this is a decision that we can make constructively.) In the first case, choosing $r \in (0, 1)$ such that $1/r < |f'(0)|$, and then using standard estimates, we obtain

$$\sup\{|f(z)| : |z| = r\} > 1.$$

Hence there exists z such that $|z| = r$ and $|f(z)| > 1$; choosing n such that $f(z) \in S_n$, we see that $a_n = 1$.

In the case where $|f'(0)| < 2$, consider any positive integer k. If $a_k = 1$, then U is the open disc with centre 0 and radius 2. It then follows from the maximal derivative property of the Riemann mapping—which holds constructively *provided that the Riemann mapping exists*—that $|f'(0)| = 2$. This contradiction implies that $a_k = 0$. Hence, in this case, $a_n = 0$ for all n.

Putting all this together, we see that the classical form of the Riemann Mapping Theorem entails LPO, and is therefore essentially non-constructive.

In order to progress from this apparent stagnation point, we use the following pathological features of our Brouwerian counterexample as a guide.

- We cannot pin down the boundary of the domain U.
- For each point $z \in D$, we cannot tell the minimum distance we need to travel from z in order to reach the outside of U; equivalently, we cannot compute the radius of the largest ball centred on z and lying inside U.

Perhaps if we were to add to the Riemann Mapping Theorem hypotheses that ensure that neither of these pathologies can occur, we would be able to recover a constructively valid form of the theorem. To this end, Cheng (1973) introduced the following notion of approximate border for a proper open subset U of $\mathbb{C}$. For simplicity, we shall only deal with the case where U is bounded.

Let z_0 be a *distinguished point* of U, and $\varepsilon > 0$. An *ε-border of U relative to z_0* is a finitely enumerable subset B of the *complement* of U,

$$\sim U \equiv \{z \in \mathbb{C} : \forall u \in U\,(z \neq u)\},$$

such that if γ is a path in $\mathbb{C}$ with left endpoint z_0, and γ keeps at least ε away from B, then γ lies in U. (A set is *finitely enumerable* if it is the range of a mapping of a finite, possibly empty, subset of $\mathbb{N}$. Roughly, a finitely enumerable set looks like a finite set but we may be unable to tell whether or not any two of its elements are distinct.)

For example, if the positive integer N is sufficiently large, then the points

$$x_k \equiv \left(1 + \frac{\varepsilon}{2}\right) \exp\left(\frac{k\pi \mathrm{i}}{N}\right) \quad (k = 1, 2, \ldots, 2N)$$

form an ε-border of D relative to 0, since the union of the open discs with centres x_k and radius ε contains the annulus

$$\left\{z : 1 - \frac{\varepsilon}{4} < |z| < 1\right\}.$$

We say that U is *mappable* if

- it is simply connected, and
- there exists $z_0 \in U$ such that for each $\varepsilon > 0$ there is an ε-border of U relative to z_0.

(It can be shown that any point of U will then serve as the distinguished point.) Thus a mappable set is one whose border is approximated arbitrarily closely by finitely enumerable subsets of the complement.

Turning to the second pathological feature of our Brouwerian counterexample, we say that U has the *maximal extent property* if there is a function $\rho : U \longrightarrow \mathbb{R}^+$ such that for each $z \in U$,

- the disc with centre z and radius $\rho(z)$ lies in U, and
- any disc with centre z and radius greater than $\rho(z)$ intersects $\sim U$.

We are now able to state the constructive *Riemann Mapping Theorem*:

> *The following are equivalent conditions on a simply connected, open, proper subset U of $\mathbb{C}$:*
>
> (i) *U is mappable;*
> (ii) *U has the maximal extent property;*
> (iii) *U is analytically equivalent to D.*

The theorem does not require that U be bounded; indeed, using a metric on $\mathbb{C}$ different from the usual one, we can cover the bounded and unbounded cases at once. For the long and difficult proof, see Chapter 5 of (Bishop and Bridges 1985).

6 The scope of constructive mathematics

I hope that, by carefully discussing the distinctive features of constructive mathematics and then illustrating them with various theorems from a relatively elementary part of analysis, I have convinced you that constructive mathematics is a viable concern, capable of producing results at levels far deeper than had been envisaged, even by some of the greatest mathematicians in our century, before the appearance of Bishop's seminal treatise. To complete that conviction, let me end by appending a list of some of the areas of mathematics that have been treated constructively over the last thirty years.

Real and complex analysis: Lebesgue measure; Picard's theorem; the Riemann Mapping Theorem.

Abstract measure and probability theory: the Daniell integral and measure spaces; the Radon–Nikodým theorem; ergodic theory; stochastic processes.

Functional analysis: the Stone–Weierstrass theorem; Hahn–Banach theorems; the Krein–Milman theorem; Banach algebras.

L_p spaces; duality in L_p and L_∞.

Hilbert space: the functional calculus and spectral theory for selfadjoint operators.

Partial Differential equations: weak solutions of the Dirichlet Problem in $\mathbb{R}^n$.

Haar measure on locally compact groups: convolution operators; the character group; the Fourier transform and duality.

Numerical mathematics: Chebyshev approximation; the Remes algorithm; approximate interpolation.

Mathematical economics: preferences and utility functions; the existence of demand functions.

Algebra: the Hilbert basis theorem; Galois theory; valuation theory; Dedekind domains.

From these create he can
Forms more real than living man,
Nurslings of immortality.

Shelley, PROMETHEUS UNBOUND.

Notes

1. When reading the proofs of this article I came across the following pertinent remark of Heyting: 'The good habit of distinguishing between results on

recursive functions obtained by intuitionistic logic and those which for their proof need classical logic is abandoned in many recent papers and books .... I regret this, because thereby the connection of the theory with the notion of effective calculability is obscured.' (Heyting 1962).

Bibliography

Berg, G., Julian, W., Mines, R., and Richman, F. (1975). The constructive Jordan curve theorem. *Rocky Mountain J. Math.*, **5**, 225–36.

Bishop, E. (1967). *Foundations of constructive analysis.* McGraw-Hill, New York.

Bishop, E. (1973). Schizophrenia in contemporary mathematics. *American Math. Soc. Colloquium Lectures.* University of Montana, Missoula.

Bishop, E. and Bridges, D. S. (1985). *Constructive analysis.* Grundlehren der Math. Wissenschaften, Vol. 279. Springer-Verlag, Heidelberg.

Bridges, D. S. (1981). A constructive proximinality property of finite-dimensional linear subspaces. *Rocky Mountain J. Math.*, **11**, 491–7.

Bridges, D. S. (1994*a*). *Computability: a mathematical sketchbook.* Graduate Texts in Mathematics, Vol. 146. Springer-Verlag, Heidelberg.

Bridges, D. S. (1994*b*). A constructive look at the real number line. In: special issue of *Synthèse* on *Real numbers: generalizations of the reals and theories of continua* (ed. P. Ehrlich), pp. 29–92.

Bridges, D. (1998). *Constructive mathematics: a foundation for computable analysis.* To appear in *J. Theoretical Computer Science.*

Bridges, D. S., and Demuth, O. (1991). Lebesgue measurability in constructive analysis. *Bull. American Math. Soc.*, **24**, 259–76.

Bridges, D. S. and Ishihara, H. (1990). Linear mappings are fairly well-behaved. *Arch. Math.*, **54**, 558–62.

Bridges, D. S. and Richman, F. (1987). *Varieties of constructive mathematics.* London Math. Soc. Lecture Notes, Vol. 97. Cambridge University Press.

Bridges, D. S., Julian, W., and Mines, R. (1989). A constructive treatment of open and unopen mapping theorems. *Zeit. Math. Logik Grundlagen Math.*, **35**, 29–43.

Bridges, D. S., Calder, A., Julian, W., Mines, R., and Richman, F. (1982). Picard's Theorem. *Trans. American Math. Soc.*, **269**, 513–20.

Brouwer, L .E .J. (1907). *Over de grondslagen der wiskunde.* Maas & van Suchtelen, Amsterdam.

Cheng, H. (1973). A constructive Riemann mapping theorem. *Pacific J. Math.* **44**, 435–54.

Dieudonné, J. (1960). *Foundations of modern analysis.* Academic Press, New York.

Dummett, M. A. E. (1977) *Elements of intuitionism.* Clarendon Press, Oxford.

Goodman, N. D. and Myhill, J. (1978). Choice implies excluded middle. *Zeit. Logik und Grundlagen der Math.* **24**, 461.

Heyting, A. (1962). After thirty years. In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (ed. E. Nagel, P. Suppes, and A. Tarski), pp. 194–7, Stanford University Press, California.

Heyting, A. (1971). *Intuitionism—an introduction* (3rd edn). North-Holland, Amsterdam.

Richman, F., Bridges, D., Calder, A., Julian, W., and Mines, R. (1982). Compactly generated Banach spaces. *Arch. Math.* **36**, 239–43.

Richman, F. (1983). Church's thesis without tears. *Journal of Symbolic Logic,* **48**, 797–803.

Richman, F. (1996). Interview with a constructive mathematician. *Modern Logic*, **6**, 247–71.

Rudin, W. (1970). *Real and complex analysis.* McGraw-Hill, New York.

Rudin, W. (1973). *Functional analysis.* McGraw-Hill, New York.

Troelstra, A. S., and van Dalen, D. (1988). *Constructivism in mathematics. An introduction*, Vols. I and II. North-Holland, Amsterdam.

van Dalen, D. (ed.) (1981). *Brouwer's Cambridge lectures on intuitionism.* Cambridge University Press.

Yessenin-Volpin, A. S. (1970). The ultra-intuitionistic criticism and the anti-traditional program for foundations of mathematics. In *Intuitionism and proof theory* (ed. J. Myhill, A. Kino, and R. Vesley), pp. 3–46. North-Holland, Amsterdam.

Department of Mathematics
University of Waikato
Hamilton
New Zealand
email: douglas@waikato.ac.nz

4

On founding the theory of algorithms

Yiannis N. Moschovakis

My topic is the problem of "founding" the theory of algorithms, part of the more general problem of "founding" computer science; whether it needs founding—which, I will argue, it does; what should count as a "foundation" for it; and why a specific "mathematical foundation" which I have proposed[1] gives a satisfactory solution to the problem—better than the most commonly accepted "standard" approach. It will be impossible to completely avoid making some comments about the general problem of "founding a mathematical discipline", but I will strive (mostly) to stay away from overly broad generalities, and concentrate on the aspects of the question which are special to algorithms.

The chapter splits naturally into two parts: a general introduction in §1–4 which lays out the problem and reviews briefly the various approaches to it in the literature, and a more specific (in some places technical) outline of the proposed solution, beginning with §5. Before anything else, however, I will start in §1 with a theorem and a proof, a simple, elementary fact which is often included in a good, first course in computer science. It will be much easier to understand what I am after by using this sample of "computer science talk" (and my slant towards it) as a starting point.

1 The mergesort algorithm

Suppose that L is a set with a fixed (total) ordering $\leq$ on it, and let L^* be the set of all *strings* (finite sequences) of members of L. A string $v = \langle v_0, \dots, v_{m-1}\rangle$ is *sorted* (in non-decreasing order), if $v_0 \leq v_1 \leq \dots \leq v_{m-1}$, and for each $u \in L^*$, $\mathrm{sort}(u)$ is the sorted "rearrangement" of u,

$$\begin{aligned}\mathrm{sort}(u) =_{\mathrm{df}} & \text{ the unique, sorted string } v \text{ such that for some permutation} \\ & \pi \text{ of } \{0, \dots, m-1\}, \quad v = \langle u_{\pi(0)}, u_{\pi(1)}, \dots, u_{\pi(m-1)}\rangle.\end{aligned}$$

The efficient computation of $\mathrm{sort}(u)$ is of paramount importance in many computing applications. Most spell-checkers, for example, view a given manuscript as a finite sequence of words and start by "alphabetizing" it, i.e., by sorting it with respect to the lexicographic ordering. The subsequent lookup of these words in the dictionary can be done very quickly, so that this initial sorting is the most critical (expensive) part of the spell-checking process.

During the preparation of this chapter, the author was partially supported by an NSF Grant.

Among the many sorting algorithms which have been studied in the literature, the **mergesort** is (perhaps) simplest to define and analyze, if not the easiest to implement. It is based on the fact that the sorting function satisfies the equation

$$\text{sort}(u) = \begin{cases} u & \text{if } |u| \leq 1, \\ \text{merge}(\text{sort}(h_1(u)), \text{sort}(h_2(u))) & \text{otherwise,} \end{cases} \tag{1.1}$$

where $|u|$ is the length of u; $h_1(u)$ and $h_2(u)$ are the first and second halves of the sequence u (appropriately adjusted when $|u|$ is odd); and the function $\text{merge}(v, w)$ is defined recursively by the equation

$$\text{merge}(v, w) = \begin{cases} w & \text{if } v = \emptyset, \\ v & \text{else, if } w = \emptyset, \\ \langle v_0 \rangle * \text{merge}(\text{tail}(v), w) & \text{else, if } v_0 \leq w_0, \\ \langle w_0 \rangle * \text{merge}(v, \text{tail}(w)) & \text{otherwise.} \end{cases} \tag{1.2}$$

Here $u * v$ is the *concatenation* operation,

$$\langle u_0, \ldots, u_{m-1} \rangle * \langle v_0, \ldots, v_{n-1} \rangle = \langle u_0, \ldots, u_{m-1}, v_0, \ldots, v_{n-1} \rangle,$$

and $\text{tail}(u)$ is the "beheading" operation on non-empty strings,

$$\text{tail}(\langle u_0, u_1, \ldots, u_{m-1} \rangle) = \langle u_1, \ldots, u_{m-1} \rangle \quad (\text{for } m > 0).$$

We establish these facts and the main property of the mergesort algorithm in four, simple propositions.

Lemma 1.1 *Equation* (1.2) *determines a unique function on strings, and such that, if v and w are sorted, then*

$$\text{merge}(v, w) = \text{sort}(v * w), \tag{1.3}$$

i.e., $\text{merge}(v, w)$ is the "merge" of v and w in this case.

Proof This is by induction on the sum $|v| + |w|$ of the lengths of the given sequences. If either $u = \emptyset$ or $v = \emptyset$, then (1.2) determines the value $\text{merge}(v, w)$ and also implies (1.3), since $\emptyset * u = u * \emptyset = u$. If both v and w are non-empty, then by the induction hypothesis

$$\text{merge}(v, \text{tail}(w)) = \text{sort}(v * \text{tail}(w)), \quad \text{merge}(\text{tail}(v), w) = \text{sort}(\text{tail}(v) * w),$$

and then (1.2) yields immediately that $\text{merge}(v, w) = \text{sort}(v * w)$, as required. □

Lemma 1.2 *For each v and w,* $\text{merge}(v, w)$ *can be computed from* (1.2) *using no more than* $|v| + |w| - 1$ *comparisons of members of L.*

Proof This is again by induction on $|v| + |w|$. At the basis, when either $v = \emptyset$ or $w = \emptyset$, (1.2) gives the value of $\text{merge}(v, w)$ using no comparisons at all. If both v and w are non-empty, then we need to compare v_0 with w_0 to determine which of the last two cases in (1.2) applies, and, then (by the induction hypothesis) we need no more than $|v| + |w| - 2$ additional comparisons to complete the computation. □

Lemma 1.3 *The sorting function* sort(u) *satisfies equation* (1.1).

Proof If $|u| \leq 1$, then u is sorted, and so $\mathrm{sort}(u) = u$, in agreement with (1.1). If $|u| \geq 2$, then the second case in (1.1) applies, and by Lemma 1.1,

$$\mathrm{merge}(\mathrm{sort}(h_1(u)), \mathrm{sort}(h_2(u))) = \mathrm{sort}(\mathrm{sort}(h_1(u)) * \mathrm{sort}(h_2(u))) = \mathrm{sort}(u),$$

as required. □

Lemma 1.4 *If* $|u| = 2^n$, *then* sort(u) *can be computed from* (1.1) *using no more than* $n \cdot 2^n$ *comparisons of members of* L.

Proof By induction on n, the result is immediate when $n = 0$, since (1.1) yields $\mathrm{sort}(u) = u$ using no comparisons when $u = \langle u_0 \rangle$ has length $2^0 = 1$. If $|u| = 2^{n+1}$, then each of the halves of u has length 2^n, and the induction hypothesis guarantees that we can compute $\mathrm{sort}(h_1(u))$ and $\mathrm{sort}(h_2(u))$ by (1.1) using no more than $(n-1) \cdot 2^{n-1}$ comparisons for each, i.e., $(n-1) \cdot 2^n$ comparisons in all. By Lemma 1.2 now, the computation of $\mathrm{merge}(\mathrm{sort}(h_1(u)), \mathrm{sort}(h_2(u)))$ can be done by (1.3) using no more than $2^n - 1 < 2^n$ additional comparisons, for a grand total of $n \cdot 2^n$. □

If we define the "binary logarithm" of a positive number by

$$\log_2(m) = \text{the least } n \text{ such that } m \leq 2^n,$$

then Lemma 1.4 (with a bit of arithmetic) yields easily the main result that we have been after:

Theorem 1.5 *The mergesort algorithm sorts a string of length* n *using no more than* $n \cdot \log_2(n)$ *comparisons.* □

The above theorem is an important result because the number of required comparisons is a very reasonable measure of complexity for a sorting algorithm, and it can be shown that $n \log_2(n)$ is *asymptotically* the least number of comparisons required to sort a string of length n.

Programming considerations The mergesort is a recursive algorithm, and so it is easiest to express in a relatively rich programming language which understands recursion, like `Pascal`, `C`, or `Lisp`—in fact, all that is needed is to re-write equations (1.1) and (1.2) in the rigid syntax of these languages;[2] it is correspondingly difficult to express it directly in the assembly language of some machine because in that case we must first implement recursion, which is not a simple matter. In addition, whether produced by the compiler of a high-level language or by hand, the implementation of the mergesort requires a good deal of space and (as with all implementations of recursive algorithms), it may be slow. Because of these reasons, the mergesort is not often used in practice, despite its simplicity and optimality.

2 Deconstruction

Before going on to learn that most of the preceding section was really meaningless gibberish, the conscientious reader should re-read it and make sure that, in fact, it makes perfect sense—except, perhaps, for the last paragraph which turned the computerese up a bit.

Lemmas 1.1 and 1.3 make straightforward, mathematical assertions about the merging and sorting functions, and their proofs are standard. Not so with Lemmas 1.2 and 1.4: they proclaim that the values of these functions *can be computed from equations* (1.2) *and* (1.1) *using no more than some number of comparisons.* Evidently, these lemmas are not just about the merging and sorting functions, but also about *computations*, *numbers of comparisons*, and (more significantly) about the specific equations (1.2) and (1.1). We understand the proof of Lemma 1.2, for example, by reading equation (1.2) as an (implicit) definition of a *computation procedure*:

2.1 The merging algorithm

To compute $\mathrm{merge}(v,w)$*, look first at* v*; if* $v = \emptyset$*, give output* w*; otherwise, if* $w = \emptyset$*, give output* v*; otherwise, if* $v_0 \leq w_0$*, compute* $z = \mathrm{merge}(\mathrm{tail}(v), w)$ *and give output* $\langle v_0 \rangle * z$*; and if none of the preceding cases applies, compute* $z = \mathrm{merge}(v, \mathrm{tail}(w))$ *and give output* $\langle w_0 \rangle * z$*.*

And here is the corresponding reading of (1.1) which we need for the proof of 1.4:

2.2 The mergesort algorithm

To sort a string u*, check first if* $|u| \leq 1$*, and if this is true, give output* u*; otherwise, sort separately the first and the second half of* u*, and then merge the values by the procedure* (2.1).

But these elaborations are not enough: We also made in the proofs of Lemmas 1.2 and 1.4 certain assumptions about the "making" and "counting" of "comparisons" by the computation procedure we extracted from equations (1.2) and (1.1). In the proof of Lemma 1.4, for example, we assumed that, *if we need* C_1 *comparisons to sort* $h_1(u)$ *and* C_2 *comparisons to sort* $h_2(u)$*, then, altogether we need* C_1+C_2 *comparisons to* (separately) *sort both of these strings.* These are very natural assumptions, to be sure, as are the interpretations of equations (1.2) and (1.1)—which is why the proofs in §1 appear to be solid. Suppose, however, that in the middle of a mathematical seminar talk about some operator $T(f)$ on Hilbert space, the lecturer appeals to the equation

$$T(f+g) = T(f) + T(g);$$

then he or she would be immediately challenged to prove that $T(f)$ is additive, starting (presumably) with a precise definition of $T(f)$, if one has not been given. What is missing in §1 are precise (mathematical) definitions of *algorithms*, *uses of comparisons*, etc., and rigorous proofs, from the definitions, of the basic properties of algorithms on which the arguments were grounded.

I have called **algorithms** these purposeful interpretations of equations (1.2) and (1.1), but *computation procedures* or *effective, deterministic instructions* could do as well (for now)—all these words are used in computer science literature, more-or-less interchangeably.

Implementations The second paragraph of §1 starts with the comment that [among sorting algorithms]

> ... the mergesort is (perhaps) simplest to define and analyze, if not the easiest to implement,

and the last paragraph of the section elaborates on the issue. Lots of new words and claims are thrown around in that paragraph. It is asserted that "the mergesort is a recursive algorithm" which can be "expressed in `Pascal` or `Lisp`"; that "it is not a simple matter to implement recursion [in an assembly language]"; that "the implementation of the mergesort requires a lot of space", etc. The innocent reader should take it on faith that all of this makes perfect common sense to an experienced programmer, and also that very little of it has ever been defined properly. Now "not the easiest" and "a lot of space" will never be made precise, to be sure, but this kind of talk suggests that programmers understand and (generally) affirm the following:

(1) A given algorithm can be expressed (programmed, implemented) in different programming languages, and so (in particular), an algorithm has many implementations.

(2) Implementations have important properties, e.g., the time and space needed for their execution.

Moral To found the theory of algorithms, we must define precisely its basic notions, starting with *algorithms*, *implementations*, and the relation between a given algorithm and its various implementations; and it is important that this be done so that the arguments in §1 are endowed with precise meaning very nearly in their present form, because these simple, intuitive ideas are so natural and appealing as to cast doubt on the necessity for rigor.

3 How do we define basic notions?

The Moral declares that we should give *precise definitions* of algorithms and implementations, but there is more than one way to go about this. Consider the following three different approaches (one with two flavors), starting with the "standard" one, which, in fact, I will adopt.

(I) Define them in set theory This is certainly the "orthodox" method of making notions precise in modern mathematics: to "found" number theory, we define the whole numbers and the operations on them in set theory; to "found" analysis, we give rigorous, set-theoretic definitions of the real numbers, functions, limits, derivatives, etc.; to "found " probability theory, we declare that "a random variable is a measurable function on a probability space," right after we give precise, set-theoretic definitions of all the words within the quotes.

Despite its wide acceptability by working mathematicians, this kind of "set-theoretic foundation" for a mathematical theory has been attacked by many philosophers, most seriously Benacerraf (1965), and also by some mathematicians; Saunders MacLane has entertained generations of audiences by asking plaintively in countless lectures:

Does anybody, *seriously* think that $2 = \{\emptyset, \{\emptyset\}\}$?

Probably not, but the von Neumann ordinal $\{\emptyset, \{\emptyset\}\}$ clearly "codes" all the properties of two-element sets which depend only on their cardinality; somewhat more fancifully, $\{\emptyset, \{\emptyset\}\}$ *models faithfully* the number 2 (whatever that is) up to *equinumerosity*—as, in fact, does any two-element set. For some less trivial examples, any *Peano system* $(M, 0, S)$ models faithfully "the natural numbers" (whatever they are), up to first-order isomorphism;[3] and any countable, dense linear ordering without endpoints models faithfully "the order type" η of the rational numbers (whatever that is), up to order-isomorphism.[4,5]

The proper role of a "set-theoretic definition" of a mathematical notion C is not to tell us in ultimate, metaphysical terms exactly *what* the C-objects (those which fall under C) *are*, but to identify and delineate their fundamental, mathematical properties. Typically, we do this by specifying a class of sets M_C and an equivalence relation $\sim_C$ on M_C, with the intention that each $\alpha \in M_C$ *faithfully represents* (codes) some C-object α_C, and that two members $\alpha, \beta \in M_C$ code the same C-object exactly when $\alpha \sim_C \beta$. A modelling of this type is successful if the $\sim_C$-invariant properties of the members of M_C capture exactly the fundamental properties of the C-objects—which implies that every fundamental property of a C-object can be "read off" any of its codes.[6]

For the case of algorithms, I will first introduce the class of *recursors*, which model the "mathematical structure of algorithms" (much like measurable functions on probability spaces model random variables), and the relation of *recursor isomorphism* between them, which models "algorithm identity". Algorithms, however, do not make sense absolutely, but only with respect to certain "data" and certain "given" (possibly higher-order) operations on these data, relative to which they are "effective"; for the full modelling, then, I will also introduce the appropriate *structures* which model such data+givens contexts (up to *structure isomorphism*), and finally claim that the recursors which are *explicitly and immediately definable* (in a specific, precise sense) on each structure $\mathfrak{M}$ model faithfully "the algorithms of $\mathfrak{M}$".

(II) Deny that they exist In the original, "naive" development of the calculus, there were real numbers, variables, limits, infinitesimals, differentials and many other things. Some of these were eventually given rigorous, set-theoretic definitions, perhaps not always completely faithful to their naive counterparts, but close enough; for example, a real-valued *function* is not exactly the same thing as a *dependent variable* and the modern notion of a *differential* is far removed from the classical one, but we can still recognize the old objects in their precise counterparts. There are, however, no *infinitesimals* in (standard) modern analysis; classical statements about infinitesimals are viewed as informal (and vague) "ways of speaking" about real numbers, functions and limits, and they must be replaced by precise statements which make no reference to them and (roughly) mean the same thing.

There are two, wildly different, approaches to the foundations of computer science which treat algorithms as "pre-mathematical" notions, to be denied rather than defined.

(IIa) Algorithms as implementations By this "standard view", especially popular among complexity theorists, there are no algorithms, only *implementations*, variously called *machines* or *models of computations*;[7] these are modeled in set theory; and assertions about algorithms like those in §1 are understood as informal "ways of speaking" about implementations. I will discuss this approach in detail in §4.

(IIb) Algorithms as constructive proofs Another, more radical proposal which also denies independent existence to algorithms is the claim that *algorithms are implicitly defined by constructive proofs.* Consider, for example, an assertion of the form

$$\phi \equiv (\forall x \in A)(\exists y \in B)P(x, y). \tag{3.1}$$

A constructive proof of ϕ should naturally yield an algorithm for computing a function $f : A \to B$, such that

$$(\forall x \in A)P(x, f(x)),$$

and there exists a considerable body of work verifying this for formalized systems of constructive mathematics, typically using various notions of *realizability* or (considerably deeper) applications of the Gentzen *cut elimination* procedure. To pursue the reduction suggested here, however, one needs to argue the converse: that statements about algorithms (in general) are really assertions about constructive proofs, and that they can be re-formulated so that all references to "algorithms" are eliminated.[8]

One problem with this view is that algorithms "support" many auxiliary notions, like "number of comparisons" and "length of computation", which are not usually associated with proofs. Girard, who is its foremost expositor, has introduced *linear logic* partly in an attempt to associate with proofs some of these notions, especially an account of *use of resources* which is often important in algorithm analysis. I suppose one could re-prove the results of §1 in some dialect of linear logic, and show that *no more than* $n \cdot \log_2(n)$ *"assumptions" of comparisons are needed to prove that* $\mathrm{sort}(u)$ *is defined*, if u has length n. This, or something very much like it, would be the assertion about constructive proofs which captures the meaning of Theorem 1.5. Now, some considerable effort is required to do this proof-theoretic analysis, and, in the end (I believe) one will again need to write down and argue from the all-important equations (1.2) and (1.1). But the *mere* (classical) *truth* of these equations suffices to "yield the algorithm" and its basic property, and so I do not see the foundational significance of constructing the linear logic proof.

Although I doubt seriously that algorithms will ever be eliminated in favor of constructive proofs (or anything else, for that matter), I think that this view is worth pursuing, because it leads to some very interesting problems. With specific, precise definitions of algorithms and constructive proofs at hand, one could investigate whether, in fact, every algorithm can be extracted (in some concrete way) from some associated, constructive proof. Results of this type would add to our understanding of the important connection between *computability* and *constructivity*.

(III) Axiomatize their theory This is what we do for set theory: Can we similarly take "algorithms", "implementations", and whatever else we need as *primitive notions* and formulate a reasonable axiomatic theory which will make sense out of computer science talk such as that in §1?

I am trying to ask a methodological question here, one which could be answered without making a commitment to any specific philosophy of mathematics. We can understand a proposed set of axioms for a theory T *formally*,[9] as being "all there is to T"; *realistically*, as expressing some important truths about the fundamental objects and notions of T, which exist independently of what we choose to say about them; and, surely, in many more, subtler ways. It seems, however, that the foundational value of a specific axiomatization (how much it helps us to understand T) is independent of our general view of the axiomatic method. It has more to do with the choice of primitives and axioms, and what the development of T from them reveals about T.[10]

I will also exclude from this option the kind of "second-order axiomatizations" which accept (uncritically, as part of logic) quantification over *all* subsets of the domain. It is often claimed, for example, that the Peano axioms provide a foundation of arithmetic in second-order logic, because of the "categoricity" theorem (b) in Note 3. This is true, as far as it goes, but we cannot account for all uses of whole numbers in mathematics by appealing to such an *external* (metamathematical) interpretation of (b). In many important applications we need to understand (b) *internally* (as part of our mathematics), for example, to prove that "every two complete, ordered fields are isomorphic".[11] This problem is even more severe for complex notions like algorithms (or topological spaces, random variables, etc.) whose basic properties are explicitly and irreducibly set-theoretic; second order "axiomatizations" can yield (at most) a poor shadow of the account of them that we need to understand their uses in mathematics.

What remains is the possibility of an axiomatization of computer science whose natural formalization would be in first-order logic, or (at least) in a many-sorted, first-order logic, where some of the basic sets are fixed to stand for *numbers* (so we can talk of "the number of comparisons" or "the number of steps" in a computation) and a few other, mathematical objects. The trouble now is that the theory is too complex: there are too many notions competing for primitive status (algorithms, implementations and computations, at the least) and the relations between them do not appear to be easily expressible in first-order terms. I doubt that the project can be carried through, and, in any case, there are no proposals on the table suggesting how we might get started.

Syntax vs. semantics Finally, I should mention—and dismiss outright—various vague suggestions in computer science literature that *algorithms are syntactic objects*, e.g., *programs*. Perhaps (Frege 1892) said it best:

> This connection [between a sign and its denotation] is arbitrary. You cannot forbid the use of an arbitrarily produced process or object as a sign for something else.

In the absence of a precise semantics, `Pascal` programs are just meaningless scribbles; to read them as algorithms, we must first interpret the language—and it is then the *meanings* attached to programs by this interpretation which are the algorithms, not the programs themselves.[12]

4 Abstract machines and implementations

The first definition of an *abstract machine* was given by Turing, in the classic (1936). Without repeating here the well-known definition (e.g., see (Kleene 1952)[13]), we recall that each *Turing machine* M is equipped with a "semi-infinite tape" which it uses both to compute and also to communicate with its environment: to determine the value $f(n)$ (if any) of the partial function[14] $f : \mathbb{N} \rightharpoonup \mathbb{N}$ computed by M, we put n on the tape in some standard way, e.g., by placing $n + 1$ consecutive 1s at its beginning; we start the machine in some specified, initial, internal state q_0 and looking at the leftmost end of the tape; and we wait until the machine stops (if it does), at which time the value $f(n)$ can be read off the tape, by counting the successive 1s at the left end. Turing argued that *the number-theoretic functions which can* (in principle) *be computed by any deterministic, physical device are exactly those which can be computed by a Turing machine*, and the corresponding version of this claim for partial functions has come to be known as the *Church–Turing Thesis*, because an equivalent claim was made by Church at about the same time. Turing's brilliant analysis of "mechanical computation" in (1936) and a huge body of work in the last sixty years has established the truth of the Church–Turing Thesis beyond reasonable doubt; it is of immense importance in the derivation of foundationally significant *undecidability results* from technical theorems about Turing machines, and it has been called "the first natural law of pure mathematics."

Turing machines capture the notion of *mechanical computability of number-theoretic functions*, by the Church–Turing Thesis, but they do not model faithfully the notion of *mechanical computation.* If, for example, we code the input by putting the argument n on the tape in *binary*[15] (rather than *unary*) notation (using no more than $\log_2(n)$ 0s and 1s), then the time needed for the computation of $f(n)$ can sometimes be considerably shortened; and if we let the machine use two tapes rather than one, then (in some cases) we may gain a quadratic speed-up of the computation; see (Maass 1985). This means that important aspects of the complexity of computations are not captured by Turing machines. We consider here a most general notion of *model of computation*, which (in particular) makes the mode of input and output part of the "machine".

Definition 4.1 *For any two sets X and W, an* **iterator** $\phi : X \rightsquigarrow W$ *is a quintuple* (input, S, σ, T, output), *where:*

(1) *S is an arbitrary (non-empty) set, the set of states of ϕ;*
(2) input : $X \to S$ *is the input function of ϕ;*
(3) *$\sigma : S \to S$ is the transition function of ϕ;*
(4) *$T \subseteq S$ is the set of terminal states of ϕ, and $s \in T \Longrightarrow \sigma(s) = s$;* and
(5) output : $T \to W$ *is the output function of ϕ.*

The **computation** *of ϕ for a given $x \in X$ is the sequence of states $\{s_n(x)\}_{n\in\mathbb{N}}$ defined recursively by*

$$\begin{aligned} s_0(x) &= \text{input}(x), \\ s_{n+1}(x) &= \begin{cases} s_n(x) & \text{if } s_n(x) \in T, \\ \sigma(s_n(x)), & \text{otherwise;} \end{cases} \end{aligned}$$

the **computation length** *on the input x (if it is finite) is*

$$\ell(x) = (\text{the least } n \text{ such that } s_n(x) \in T) + 1;$$

and the partial function $\overline{\phi} : X \rightharpoonup W$ **computed by** *ϕ is defined by the formula*

$$\overline{\phi}(x) = \text{output}(s_{\ell(x)}(x)).$$

Each Turing machine M can be viewed as an iterator $M : \mathbb{N} \rightsquigarrow \mathbb{N}$, by taking for states the (so-called) "complete configurations" of M, i.e., the triples (σ, q, i) where σ is the tape, q is the internal state, and i is the location of the machine, along with the standard input and output functions.

It is generally conceded that this broad notion of iterator can model the manner in which every conceivable (deterministic, discrete, digital) mechanical device computes a function, and so it captures the *structure of mechanical computation.* It is too wide to capture the *effectivity of mechanical computation*, because it allows an arbitrary set of states and arbitrary input, transition and output functions, but (for the moment) I will disregard this problem; it is easy enough to solve by imposing definability or finiteness assumptions on the components of iterators, similar to those of Turing machines, see Proposal IV in §8. The question I want to address now is whether the notion of iterator is *wide enough* to model faithfully *algorithms*, as it is typically assumed in complexity theory;^{16} put another way,

$$\text{are algorithms the same as mechanical computation procedures?} \qquad (4.1)$$

A positive answer to this question expresses more precisely the view (**IIa**) in §3 and it might appear that it is the correct answer, especially as we have been using the two terms synonymously up until now. There are, however, at least two serious problems with this position.

Recursion and iteration If all algorithms are modeled by iterators, then which iterator models the mergesort algorithm of §1? This was defined implicitly by the *recursive equations* (1.1) and (1.2) (or so we claimed in §1), and so we first need to transform the intuitive computation procedure which we extracted from these equations into a precise definition of an iterator. The problem is not special to the mergesort, which is just one of many important examples of *recursive algorithms* defined by systems of recursive equations.

To clarify the situation, consider the following description of an arbitrary iterator $\phi = (\text{input}, S, \sigma, T, \text{output})$ by a *while-program* in a pidgin, Pascal-like programming language:

```
$s := \text{input}(x)$;
while ($s \notin T$) $s := \sigma(s)$;
$w := \text{output}(s)$;
return $w$.
```

We do not need any elaborate, precise definitions of the semantics of while-programs to recognize that this one (naturally understood) defines ϕ, and that, conversely, the algorithm expressed by any program built with assignments and while-loops can be directly modeled by an iterator. The first problem, then, is how to construct while-programs which express the intuitive computation procedures implicit in systems of recursive equations like (1.1) and (1.2).

This can be done, in many different ways generally called *implementations of recursion.*[17] These methods are not simple, but they are precise enough so that they can be automated. For example, one of the most important tasks of a *compiler* for a "higher level" language like Pascal is exactly this conversion of *recursive programs* to *while-programs*, in the assembly language of a specific processor (a concrete, physical iterator, really), which can then run them.

Assume then that we associate with each system E of recursive equations (like (1.1) and (1.2)) an iterator ϕ_E, using some fixed "compilation process", and we make the view (**IIa**) precise by calling ϕ_E *the algorithm defined by E.* Now the **first problem** with this view is that ϕ_E is far removed from E and the resulting rigorous proofs of the important properties of ϕ_E are complex and only tenuously related to the simple, intuitive arguments outlined in §1.

The complaint is not so much about the mere complexity of the rigorous proofs, because it is not unusual for technical complications to crop up when we insist on full rigor in mathematics. It is the artificiality and irrelevance of many of the necessary arguments which casts doubt on the approach, as they deal mostly with the specifics of the compilation procedure rather than the salient, mathematical properties of algorithms. Still, this is not a fatal objection to (**IIa**), only an argument against it, on the grounds that the loss of elegance and simplicity which it requires is out of proportion with the gain in rigor that it yields.

The non-uniqueness of compilation The **second problem** with the view (**IIa**) is that there are many ways to "compile" recursive programs—to assign an iterator ϕ_E to each system of recursive equations E—and there is no single, natural way to choose any one of them as "canonical". This is a most serious problem, I think, which makes it very unlikely that we can usefully identify algorithms with computational procedures, or iterators.

Take the mergesort, for example, express it formally in `Pascal`, `C` and `Lisp`, and suppose ϕ_P, ϕ_C and ϕ_L are the iterators which we get when we compile these programs in some specific way for some specific processor. Each of these three iterators has equal claim to be "the mergesort algorithm" by (**IIa**), and there is no obvious way to choose among them. More significantly (because we might allow ourselves some arbitrary choice here), these three iterators, obviously, have something in common, but exactly

$$\text{what is the relation between } \phi_P,\ \phi_C \text{ and } \phi_L? \tag{4.2}$$

The natural answer is that

$$\text{they are all implementations of the mergesort algorithm,} \tag{4.3}$$

but, of course, we cannot say this without an independent notion of *the mergesort algorithm.* Even if we give up on making precise and answering fully Question (4.2), we would still like to say that

> every computational procedure extracted from the recursive equations (1.1) and (1.2) satisfies Lemmas 1.2 and 1.4

(suitably formulated for iterators), and it is hard to see how we can express this without making reference to some one, semantic object, assigned directly to (1.1) and (1.2) and with a prior claim to model *the mergesort algorithm.*

Proposal I: Implementations are iterators From this discussion, it seems to me most natural to assume that *iterators model implementations*, which are special, "iterative algorithms," and that results such as Lemmas 1.2 and 1.4 are about more abstract objects, whatever we decide to call them; each of these objects, then, may admit many implementations, and codes the "implementation independent" properties of algorithms.

5 The theory of recursive equations

To motivate our choice of set-theoretic representations of algorithms in the next section, let us first outline rigorous formulations and proofs of the results in §1 in the context of the *theory of recursive equations.* This is a simple, classical theory, whose basic results are very similar in flavor to those of the *theory of differential equations.*

A *poset* (partially ordered set)[18] $(D, \leq_D)$ is *inductive* or *complete* if every chain (linearly ordered subset) $A \subseteq D$ has a least upper bound, $\sup A$, and a mapping (function)

$$\pi : D \to E$$

on one poset to another is *monotone* if

$$d \leq_D d' \implies \pi(d) \leq_E \pi(d').$$

The basic fact about complete posets is that monotone mappings have least fixed points, in the following, strong sense.

Theorem 5.1 (The monotone, least fixed point theorem) *Suppose that*

$$\pi : X \times D \to D$$

is a monotone mapping on the poset product $X \times D$ to D, and that D is inductive. Then, for each $x \in X$, the equation

$$d = \pi(x, d) \quad (x \in X, d \in D)$$

has a least solution

$$d(x) = (\mu d \in D)[d = \pi(x, d)],$$

characterized by the conditions that

$$d(x) = \pi(x, d(x)), \quad (\forall e \in D)[e \leq_D \pi(x, e) \Longrightarrow d(x) \leq_D e];$$

in addition, the function $x \mapsto d(x)$ is monotone on X to D.[19] □

The simplest, interesting inductive posets are the partial function spaces

$$(A \rightharpoonup B) = \{p \mid p : A \rightharpoonup B\} \quad (= \{p \mid p : A \to B \cup \{\bot\}\})$$

partially ordered "pointwise",

$$\begin{aligned} p \leq q &\iff (\forall x \in A)[p(x) \leq q(x)] \\ &\iff (\forall x \in A, y \in B)[p(x) = y \Longrightarrow q(x) = y], \end{aligned}$$

and products of these, i.e., spaces of pairs (or tuples) of partial functions. To apply Theorem 5.1 to the sorting problem of §1, for example, we need the posets $(L^* \rightharpoonup L^*)$ and $(L^* \times L^* \rightharpoonup L^*)$, which contain the functions sort and merge, and also the poset

$$(L \times L \rightharpoonup \{\mathit{ff}, \mathit{tt}\}),$$

where $\{\mathit{ff}, \mathit{tt}\}$ is some arbitrary set of two, distinct objects standing for *falsity* and *truth* and which contains the *characteristic function*

$$\chi_{\leq}(s,t) = \begin{cases} \mathit{tt} & \text{if } s \leq t, \\ \mathit{ff} & \text{if } t < s, \end{cases}$$

of the given ordering on L. In general, a partial function $c : L \times L \rightharpoonup \{\mathit{ff}, \mathit{tt}\}$ can be viewed as the *characteristic partial function*, of a *partial, binary relation* on L. The idea is to generalize the problem, and try to find (partial) "merging" and "sorting" functions, relative to an arbitrary partial relation $c : L \times L \rightharpoonup \{\mathit{ff}, \mathit{tt}\}$, which stands for some approximation to a total ordering. We can get this very easily from Theorem 5.1: *for each $c : L \times L \rightharpoonup \{\mathit{ff}, \mathit{tt}\}$, there exist partial functions*

$$\mathrm{sort}(c) : L^* \rightharpoonup L^* \quad \text{and} \quad \mathrm{merge}(c) : L^* \times L^* \rightharpoonup L^*,$$

which are (least) *solutions of the recursive equations*

$$\mathrm{sort}(c)(u) = \begin{cases} u & \text{if } |u| \leq 1, \\ \mathrm{merge}(c)(\mathrm{sort}(c)(h_1(u)), \mathrm{sort}(c)(h_2(u))) & \text{otherwise,} \end{cases} \qquad (5.1)$$

$$\mathrm{merge}(c)(v,w) = \begin{cases} w & \text{if } v = \emptyset, \\ v & \text{else, if } w = \emptyset, \\ \langle v_0 \rangle * \mathrm{merge}(c)(\mathrm{tail}(v), w) & \text{else, if } c(v_0, w_0) = tt, \\ \langle w_0 \rangle * \mathrm{merge}(c)(v, \mathrm{tail}(w)) & \text{else, if } c(v_0, w_0) = f\!f; \end{cases} \tag{5.2}$$

and which depend monotonically on $c : L \times L \rightharpoonup \{f\!f, tt\}$. If $\leq$ is the given ordering on L, then merge($\chi_{\leq}$) and sort($\chi_{\leq}$) are obviously the merging and sorting functions we need; and on the other hand, using *exactly* the arguments (by induction on $|v| + |w|$ and $|u|$) for Lemmas 1.2 and 1.4, we can show the following:

Theorem 5.2 *Suppose that* sort(c) *and* merge(c) *are monotonic functions of* $c : L \times L \rightharpoonup \{f\!f, tt\}$ *which satisfy the recursive equations* (5.1) *and* (5.2).

(a) *If* $\mathrm{merge}(c)(v,w) = z \in L^*$, *then there exists a partial function* $c' \leq c$ *which is defined on at most* $|v|+|w|-1$ *pairs, and such that* $\mathrm{merge}(c')(v,w) = z$.

(b) *If* $|u| = 2^n$ *and* $\mathrm{sort}(c)(u) = z \in L^*$, *then there exists a partial function* $c' \leq c$ *which is defined on at most* $n \cdot 2^n$ *pairs, and such that* $\mathrm{sort}(c')(u) = z$. □

There is no mention of "algorithms" or "uses of comparisons" in Theorem 5.2, but it is not hard to find in it the heart of the claims of Lemmas 1.2 and 1.4. The key move is from the equations (1.1) and (1.2) (which *we know* to hold of the sorting and merging functions), to the "parametrized" equations (5.1), (5.2), whose meaning is unclear for arbitrary c, but which have least solutions sort(c) and merge(c) by Theorem 5.1, and these solutions depend monotonically on "the parameter" c. Let us now make the natural assumption that any method for extracting a computation procedure (perhaps an iterator) ϕ from the equations (1.1) and (1.2), should also apply to and yield a generalized computation procedure $\phi(c)$, for each c, which computes sort(c)—simply by replacing each instruction to *check if* $s \leq t$ by *compute* $c(s,t)$. If $\mathrm{sort}(u) = z$, so that ϕ applied to u computes z, then $\mathrm{sort}(c)(u) = z$, for some small $c \leq \chi_{\leq}$ by Theorem 5.2, and hence $\phi(c)$ applied to u should also compute z—but it cannot "ask" for comparisons outside the domain of c, because then it would diverge.

This simple method of *varying the parameter* (here the ordering $\leq$ on L) and then applying Theorem 5.1, is a powerful tool for deriving properties of functions which are (least) solutions of recursive equations.

6 Functionals and recursors

What do we learn from the rigorous arguments of the preceding section about choosing a set-theoretic object to model "the mergesort algorithm"? It seems that all we needed was the "semantic content" of equations (1.1) and (1.2), i.e., the pair (f,g) of operations defined by their right-hand-sides,

$$f(u,p,q) = \begin{cases} u & \text{if } |u| \leq 1, \\ q(p(h_1(u)), p(h_2(u))) & \text{otherwise,} \end{cases} \tag{6.1}$$

$$g(v,w,p,q) = \begin{cases} w & \text{if } v = \emptyset, \\ v & \text{else, if } w = \emptyset, \\ \langle v_0 \rangle * q(\text{tail}(v), w) & \text{else, if } v_0 \le w_0, \\ \langle w_0 \rangle * q(v, \text{tail}(w)) & \text{otherwise.} \end{cases} \tag{6.2}$$

Formally, these are *functionals on L^**, in a technical sense which is basic and useful enough to deserve special billing.

Definition 6.1 *A* **functional** *on a collection of sets $\mathcal{M}$ is any monotone, partial function*

$$h : X_1 \times \cdots \times X_n \rightharpoonup W,$$

where $W \in \mathcal{M}$ or $W = \{\mathit{ff}, \mathit{tt}\}$; and each X_i is either a set in $\mathcal{M}$, or a partial function space $X_i = (U \rightharpoonup V)$, with $U = U_1 \times \cdots \times U_l$ a product of sets in $\mathcal{M}$ and $V \in \mathcal{M}$ or $V = \{\mathit{ff}, \mathit{tt}\}$.

For example, the operation of *m-ary partial function application*

$$\text{app}_m(x_1, \ldots, x_m, p) = p(x_1, \ldots, x_m) \quad (x_1, \ldots, x_m \in M, p : M^m \rightharpoonup W) \tag{6.3}$$

is a functional on the sets M, W; and the operation

$$\exists_M(p) = \begin{cases} \mathit{tt} \ \text{ if } (\exists x \in M)[p(x) = \mathit{tt}], \\ \mathit{ff} \ \text{ if } (\forall x \in M)[p(x) = \mathit{ff}], \end{cases} \tag{6.4}$$

is a functional on M which "embodies" (in Kleene's expression) existential quantification on M. Note also that, by this definition, all partial functions and partial relations on M are functionals.

It was (essentially) *systems of functionals* like (f, g) that I chose initially in (Moschovakis 1984, 1989*b*) to model algorithms, and these are the concrete objects which come up in the most interesting applications. To develop the general theory simply and smoothly, however, it is best to use a class of more abstract objects, which includes suitable representations of these systems.[20]

Definition 6.2 *A* **recursor** $\alpha : X \rightsquigarrow W$ *on a poset X (perhaps discrete, just a set) to a set W is a triple (D, τ, value), where:*

(1) D is an inductive poset, the domain or solution set of α;

(2) $\tau : X \times D \to D$ is a monotone mapping, the transition mapping of α; and

(3) value $: X \times D \rightharpoonup W$ *is a monotone, partial mapping, the value mapping of α.*[21]

The partial function $\overline{\alpha} : X \rightharpoonup W$ determined (computed) by α is defined by

$$\overline{\alpha}(x) = \text{value}(x, (\mu d \in D)[d = \tau(x, d)]),$$

where, for each $x \in X$, $(\mu d \in D)[d = \tau(x, d)]$ is the least, fixed point of the recursive equation

$$d = \tau(x, d) \quad (x \in X, d \in D);$$

and it is monotone, by Theorem 5.1. We say that α is a **recursor on** *a collection of sets $\mathcal{M}$ if $\overline{\alpha} : X \rightharpoonup W$ is a functional on $\mathcal{M}$ as in Definition 6.1.*

Two recursors $\alpha_1 = (D_1, \tau_1, \text{value}_1), \alpha_2 = (D_2, \tau_2, \text{value}_2) : X \rightsquigarrow W$ *(on the same* X *and* W*) are* **isomorphic** *if there exists an order-preserving bijection*

$$\pi : D_1 \rightarrow D_2$$

which respects the transition and value mappings, i.e., for all $x \in X$ *and* $d \in D_1$,

$$\begin{aligned} \pi(\tau_1(x,d)) &= \tau_2(x, \pi(d)), \\ \text{value}_1(x,d) &= \text{value}_2(x, \pi(d)). \end{aligned}$$

Isomorphic recursors (easily) determine the same partial functions, that is $\overline{\alpha}_1 = \overline{\alpha}_2$.

Proposal II : Algorithms are recursors *The mathematical structure of every algorithm on a poset* X*to a set* W *is modelled faithfully by some recursor* $\alpha : X \rightsquigarrow W$; *and two recursors model the same algorithm if they are isomorphic.*

The where notation Defining and manipulating recursors becomes much easier with the following, compact where notation, one of several variants of the notation for recursive definitions used in programming languages: to specify that $\alpha = (D, \tau, \text{value}) : X \rightsquigarrow W$, we write

$$\alpha(x) = \text{value}(x, d) \text{ where } \{d = \tau(x, d)\}, \tag{6.5}$$

suggesting that to compute the value $\overline{\alpha}(x)$ using α, we first take the least solution of the equation within the braces { } and then plug it into the "head" partial mapping in the front. We can have more than one equation within the braces in this notation,

$$\begin{aligned} \alpha(x) = \ & \text{value}(x, d_1, d_2) \text{ where } \{d_1 = \tau_1(x, d_1, d_2), d_2 = \tau_2(x, d_1, d_2)\} \\ =_{\text{df}} \ & \text{value}(x, \langle d_1, d_2\rangle) \text{ where } \{\langle d_1, d_2\rangle = \langle \tau_1(x, d_1, d_2), \tau_2(x, d_1, d_2\rangle\}, \end{aligned}$$

where the angled brackets indicate that the domain of α is the product poset $D_1 \times D_2$; and we can also allow recursive equations involving (partial) functions within the braces,

$$\begin{aligned} \alpha(x) = \ & \text{value}(x, p) \text{ where } \{p(u) = \tau(x, u, p)\} \\ =_{\text{df}} \ & \text{value}(x, p) \text{ where } \{p = \lambda(u)\tau(x, u, p)\}, \end{aligned}$$

in which case the domain of α is the partial function poset $(U \rightharpoonup W)$, the range of the variable p.[22]

The judicious application and combination of these conventions facilitates significantly the definition and manipulation of recursors. For example, each monotone partial function $f : X \rightharpoonup W$ (and, in particular, each functional) is naturally represented by the "degenerate" recursor

$$\mathbf{r}_f(x) =_{\text{df}} f(x) \text{ where } \{d = d\},$$

with domain $\{\bot\}$ and such that (obviously) $\overline{\mathbf{r}}_f = f$. Less trivially, each iterator $\phi = (\text{input}, S, \sigma, T, \text{output})$ on X to W, is represented by the recursor

$$\mathbf{r}_\phi(x) =_{\text{df}} p(\text{input}(x)) \text{ where } \{p(s) = \text{if } s \in T \text{ then output}(s) \text{ else } p(\sigma(s))\} \quad (6.6)$$

with domain the partial function poset $(S \rightharpoonup W)$, which computes the same partial function $\overline{\mathbf{r}}_\phi(x) = \overline{\phi}(x)$ as ϕ and codes ϕ up to iterator isomorphism.[23] Finally, the "systems of functionals" which arise in the study of recursive equations can also be represented by recursors, e.g., we set

$$\text{mergesort}_1(u) = p(u) \text{ where } \{p(u) = f(u,p,q), q(v,w) = g(v,w,p,q)\}, \quad (6.7)$$

where f and g are defined by (6.1), (6.2).[24]

It is natural and convenient to "identify" monotone partial functions, iterators and systems of functionals with these recursors which represent them, so that the class of recursors may be said to include these objects.

Algorithm identity Suppose that A is an (intuitive) algorithm which computes (say) the first one billion prime numbers, and you define A' by saying

> *first add* $2+2$ *and then do* A;

or, you let A'' be

> *do* A *two times* (simultaneously, in parallel) *and give as output just one of the results*:

Are A, A' and A'' *different* algorithms, or are they all *identical*? They are, clearly, very closely related, but most people would call them different—or grant, at least, that any rigorous representation of algorithms would model them by non-isomorphic objects; and, indeed, if α, α' and α'' are their natural recursor representations, then no two of these three recursors are isomorphic.

In fact, recursor isomorphism is a very fine equivalence relationship which is not preserved by many useful algorithm transformations (optimizations, refinements, etc.), and we must often introduce "coarser" equivalences (or reductions) to express important facts of informal algorithm analysis.[25] Rather than a defect, this is a virtue, in most cases, as it forces out a precise version of exactly what it is which is being proved.

Infinitary recursors; graph connectivity It is clear that not every recursor models an algorithm,[26] because we have allowed the transition and value mappings to be completely arbitrary, as they are for iterators. We will deal with this question of "effective definability" of algorithms in §8. In contrast to iterators, however, a recursor may fail to determine an "effective computation" in a more dramatic way, as in the following example.

Suppose that (G, R) is a *graph*, i.e., a non-empty set of *nodes* G together with a symmetric, binary *edge relation* R on G, and consider the *recursive equivalence*

$$p(x,y) \iff x = y \vee (\exists z)[R(x,z) \,\&\, p(z,y)] \quad (p \subseteq G \times G). \qquad (6.8)$$

Quite easily, the least binary relation $\overline{p}$ on G which satisfies (6.8) is

$$\overline{p}(x,y) \iff \text{there is a path which joins } x \text{ with } y, \tag{6.9}$$

and from this it follows that, if we set

$$\begin{aligned}\mathbf{conn} \Leftrightarrow (\exists x)q(x) \text{ where } \{&q(x) \Leftrightarrow (\forall y)p(x,y),\\ &p(x,y) \Leftrightarrow x = y \vee (\exists z)[R(x,z)\,\&\,p(z,y)]\},\end{aligned} \tag{6.10}$$

then $\mathbf{conn} : \mathbf{I} \rightsquigarrow \{\mathit{ff}, \mathit{tt}\}$ is a nullary[27] recursor which "verifies" the connectedness of the graph G, i.e.,

$$\overline{\mathbf{conn}} \Leftrightarrow \mathit{tt} \iff G \text{ is connected.}$$

We can also extract from the recursive equivalence (6.8) a computation procedure (of sorts) for verifying whether an arbitrary $x \in G$ can be joined with some arbitrary $y \in G$, much as we did in §1: *If $x = y$, give output tt, and if not, check* (simultaneously) *for each immediate neighbour z of x, if it can be joined with y, and give tt only if one does.* So far, so good, but *how long—how many basic, computation steps—does it take to verify that G is connected*, i.e., to carry out all the verifications required to show that *every x* can be joined with *every y* in G? Well, it depends on the so-called *diameter* of G, the supremum of shortest paths connecting its points. If this is finite (and, in particular, if G is finite), then we can clearly do all the verifications in a finite number of steps, but if G is connected with infinite diameter, then it seems that we need to use infinitely many steps to check that every point in G can be joined with every other one, and so the total "computation" of $\overline{\mathbf{conn}}$ requires at least ω (= the least infinite ordinal) steps.

Whether (in the proper context) we can take **conn** to represent an "algorithm" is an interesting question, to which I will return in §9—but, if it does, then that should be some sort of (absolutely) non-implementable, *infinitary* algorithm, since "real," terminating computations cannot take infinitely many steps for their completion.

The "number of steps" required by a recursor α to "compute" a value $\overline{\alpha}(x)$ is an important quantity associated with α, part of a bundle of notions with which the mathematical theory of recursors starts.

Recursor iteration Fix a recursor $\alpha = (D, \tau, \text{value}) : X \rightsquigarrow W$ and some $x \in X$, let

$$\tau_x(d) = \tau(x,d),$$

and for each ordinal number ξ, set (by ordinal recursion)

$$\begin{aligned} d^\xi_\alpha(x) &=_{\text{df}} \tau_x(\sup\{d^\eta_\alpha(x) \mid \eta < \xi\}) \quad (\text{with } \sup\emptyset = \bot),\\ \overline{\alpha}^\xi(x) &=_{\text{df}} \text{value}(x, d^\xi_\alpha(x)),\\ ||\alpha|| &=_{\text{df}} \text{the least } \xi\ (\forall x \in X)[d^\xi_\alpha(x) = \sup\{d^\eta_\alpha(x) \mid \eta < \xi\}]. \end{aligned} \tag{6.11}$$

It is not hard to show that these definitions make sense[28] and that they determine the partial function computed by α, i.e.,

$$\overline{\alpha}(x) = \sup_{\xi}\overline{\alpha}^{\xi}(x).$$

We call α **finitary** if $||\alpha|| \leq \omega$, and **infinitary** if $||\alpha|| > \omega$.

The **closure ordinal** $||\alpha||$ and the (partial) **stage assignment**

$$|\alpha|(x) =_{\mathrm{df}} \mu\xi\ [\overline{\alpha}^{\xi}(x) \in W] < ||\alpha||, \tag{6.12}$$

(defined exactly when $\overline{\alpha}(x)$ is defined) are fundamental invariants of α: for the recursor $\mathbf{r}_{\phi}$ associated with an iterator ϕ by (6.6), for example, $||\mathbf{r}_{\phi}|| = \omega$, and

$$|\mathbf{r}_{\phi}|(x) = \ell(x) - 1 = (\text{the computation length on } x) - 1.$$

In the case of 6.6,

$$||\mathbf{conn}|| = \text{the diameter of } G + 2,$$

so that if G has infinite diameter, then $||\mathbf{conn}|| = \omega + 2$.

One may choose to view the iteration sequence $\{d^{\xi}(x) \mid \xi < ||\alpha||\}$ as some sort of very abstract, "logical computation" of $\overline{\alpha}(x)$, whose length (if it terminates) is the possibly infinite ordinal $|\alpha|(x)$. More loosely, but closer to the truth, we may say that each iterate $d^{\xi}(x)$ codes some "information" about the value $\overline{\alpha}(x)$, which can be extracted by the value mapping and increases with ξ; and when enough such information is available, then $\overline{\alpha}(x) = \overline{\alpha}^{\xi}(x) = \text{value}(x, d^{\xi}(x))$ becomes defined.

These iterates are also the key tool for "rigorizing" many informal arguments about algorithms extracted from recursive equations. I will not go into this here, and I will also avoid any further discussion of the mathematical theory of recursors, whose basic facts are presented in (Moschovakis 1989*a*, 1989*b*, 1994*b*).

7 Implementations

Imagine a world (presumably run by mathematicians) where one could patent algorithms, so that each time you used Professor X's "superfast multiplication" α you should pay him a fee.[29] Now to use α, you must first *implement* it, i.e., write a program P which (in some sense) expresses α, and which can be understood and "run" by some actual machine; and Professor Y has written just such a program P, but he claims that it has nothing to do with X's α, it is actually an implementation of some other algorithm β, unrelated to α and invented by himself. What are the relevant objective criteria—the mathematical relations which hold or do not hold between α and P or β and P—for settling the dispute? The humor is dubious, but the problem of making precise exactly what it means to say that *the program P implements the algorithm α* is very serious, one of the basic (I think) foundational problems in the theory of algorithms.

Having already resolved in **Proposal I** that implementations are iterators, and that each iterator ϕ can be identified with a recursor $\mathbf{r}_{\phi}$, (6.6), I will propose an answer which follows from a general theory of *reduction* among recursors. First I will define a relation $\alpha \leq_r \beta$ between recursors, which (roughly) means that "the abstract computations of α are faithfully simulated by those of β", and then I will say that ϕ *implements* α if $\alpha \leq_r \mathbf{r}_\phi$.

Definition 7.1 *A recursor* $\alpha_1 = (D_1, \tau_1, \text{value}_1) : X_1 \rightsquigarrow W$ *is* **reducible** *to another* $\alpha_2 = (D_2, \tau_2, \text{value}_2) : X_2 \rightsquigarrow W$ (on the same set of values), *and we write* $\alpha_1 \leq_r \alpha_2$, *if there exist monotone mappings*

$$\rho : X_1 \to X_2, \quad \pi : X_1 \times D_1 \to D_2,$$

so that:

(1) *For all* $x \in X_1$ *and* $d \in D_1$, $\tau_2(\rho(x), \pi(x,d)) \leq \pi(x, \tau_1(x,d))$;

(2) *for all* $x \in X_1$ *and* $d \in D_1$, $\text{value}_2(\rho(x), \pi(x,d)) \leq \text{value}_1(x,d)$*;* and

(3) *for each* $x \in X_1$, $\overline{\alpha}_1(x) = \overline{\alpha}_2(\rho(x))$.

It is easy to show that the reduction relation is reflexive and transitive on the class of all recursors.

Proposal III: "To implement" means "to reduce" *An* **implementation of** *a recursor* α *is any iterator* ϕ *such that* $\alpha \leq_r \mathbf{r}_\phi$*; and* α *is* (*abstractly*) **implementable** *if it admits an implementation.*

In the concrete examples of this very abstract notion, the universe X_2 of α_2 is an expansion of the universe X_1 of α_1 by new "data structures," e.g., stacks and caches. To understand how the abstract computations of the two recursors are related, imagine (as at the end of §6) that each $d \in D_1$ represents some information about the value $\overline{\alpha}_1(x)$, which by (3) of Definition 7.1 is the same as $\overline{\alpha}_2(\rho(x))$; for each x, now, $\pi(x,d)$ gives us a corresponding piece of information about $\overline{\alpha}_2(\rho(x))$, and (1) and (2) prescribe that each step of α_2 "increments no more" the available information and "contributes no more" to the computation of the common value $\overline{\alpha}_1(x) = \overline{\alpha}_2(\rho(x))$ than the corresponding step of α_1.[30] Technically, (1) and (2) yield that for all ordinals ξ,

$$\pi(x, d_1^\xi(x)) \geq d_2^\xi(\rho(x)),$$

from which it follows that

$$\overline{\alpha}_1^\xi(x) \geq \overline{\alpha}_2^\xi(\rho(x)), \tag{7.1}$$

and (3), then, says simply that the iteration of α_2 eventually "catches up" with that of α_1, so that, in the limit, the same partial function is computed.

From (7.1) we also obtain

$$|\alpha_1|(x) \leq |\alpha_2|(\rho(x)) \quad (\text{if } \overline{\alpha}_1(x) \text{ is defined}),$$

so that in particular

$$||\alpha_1|| \leq ||\alpha_2||; \tag{7.2}$$

and this implies that, *if* α *is abstractly implementable, then* $||\alpha|| \leq \omega$. It is not hard to verify that the converse is also true,[31] so that the following holds.

Proposition 7.2 *The abstractly implementable recursors are exactly those with closure ordinal $\leq \omega$.* □

To justify this modelling of "ϕ implements α", one must (at least) show that it covers simply and naturally the standard reductions of recursion to iteration, and that it extends the precise definitions which already exist for simulating one iterator by another. This can be done, quite easily, but the inevitable technicalities are not suitable for this paper.

8 Algorithms

It is tempting to assume that the successor operation $S(n) = n+1$ on the natural numbers is "immediately computable," an absolute "given," presumably because of the trivial nature of the algorithm for constructing the unary (tally) representation of $S(n)$ from that of n—just *add one tally*; if we use binary notation, however, then the computation of $S(n)$ is not so trivial, and may require the examination of all $\log_2(n)$ binary digits of n for its construction—while multiplication by 2 becomes trivial now—*just add one* 0. This point was made in §4, to argue that the mode of input must be part of any model of computation, but it also shows that while there is one absolute notion of *computability* on $\mathbb{N}$ (by the Church–Turing Thesis), there is no corresponding absolute notion of "algorithm" on the natural numbers—much less on arbitrary sets. Algorithms make sense only *relative* to operations which we wish to admit as *immediately given* on the relevant sets of *data*. Any set can be a data set; as for the given operations, we may have partial functions, functionals, or, in the most general case, recursors.

Definition 8.1 *A* (recursor) **structure** *is a pair $\mathfrak{M} = (\mathcal{M}, \mathcal{F})$ such that the following hold.*

(1) *Each $M \in \mathcal{M}$ is a set and at least one such M is non-empty; these are the* **basic** *or* **data** *sets of $\mathfrak{M}$.*

(2) *Each $\alpha \in \mathcal{F}$ is a recursor on the data sets of $\mathfrak{M}$, i.e., $\overline{\alpha} : X \rightharpoonup W$ is a functional on $\mathcal{M}$ as in* 6.1.

$\mathfrak{M}$ is a first-order structure if every given is a (total) *function, and a functional structure if every given is a functional.*

Simplest among these are the usual (single- or many-sorted) first-order structures of model theory, e.g.,

$$\mathfrak{N} = (\mathbb{N}, 0, \mathbb{S}, \mathbb{P}, \chi_0), \tag{8.1}$$

where 0 is the (nullary) constant; $S(n) = n+1$; $P(0) = 0$ and $P(n+1) = n$; and $\chi_0 : \mathbb{N} \to \{ff, tt\}$ is the characteristic function of $\{0\}$; the choice of givens in this simplest *structure of arithmetic* implies (in effect) that we take the numbers to be finite sequences of tallies. The expansion

$$(\mathfrak{N}, \exists_{\mathbb{N}}) = (\mathbb{N}, 0, \mathbb{S}, \mathbb{P}, \chi_0, \exists_{\mathbb{N}}), \tag{8.2}$$

Term:	$A :\equiv \mathit{ff}$	$\mid \mathit{tt}$
		$\mid p(Z_1,\dots,Z_n)$
		$\mid \mathsf{f}(Z_1,\dots,Z_n,\pi_1,\dots,\pi_m)$
		$\mid$ if A_0 then A_1 else A_2 fi
(for FLR)		$\mid A_0$ where $\{p_1(\vec{u}_1) = A_1,\dots,p_n(\vec{u}_n) = A_n\}$
(for FLI)		$\mid p(\vec{I})$ where $\{p(\vec{s}) =$ if T then O else $p(\vec{Z})$ fi$\}$
	$Z :\equiv u$	$\mid A$
λ-term:	$\pi :\equiv p$	$\mid \lambda(u_1,\dots,u_n)A$

TABLE 1. The syntax of FLR and FLI.

of $\mathfrak{N}$ by the existential quantifier (6.4) is an important example of a functional structure, and every first-order structure $\mathfrak{M}$ has an analogous expansion $(\mathfrak{M}, \exists_{\mathfrak{M}})$. In the most general case, the "given" recursors of a structure represent algorithms which we accept as "black boxes," right from the start, and they are the basic ingredients with which we build the algorithms of a structure.

Formal definability With each structure $\mathfrak{M} = (\mathcal{M}, \mathcal{F})$, we can associate a *vocabulary* (signature), with variables over the data sets of $\mathfrak{M}$ and the partial functions and partial relations on them, and with constant, function symbols for the given recursors; and using such a vocabulary, we can then build formal languages, which codify various notions of definability on $\mathfrak{M}$. Simplest among these is the *Formal Language of Recursion* FLR, a language of terms and λ-terms, built up from the vocabulary using (partial) *function application* (6.3), *definition by cases* (conditional), *calls* to the givens, and *recursion*, using the "where" notation of §6. We will also need the fragment FLI of FLR, obtained by replacing the general where construct by its special case used in iteration. Table 1 gives a summary definition of the syntax of these languages, computer science style.

The **algorithmic** or **intensional semantics** of FLR which concern us here are defined in (Moschovakis 1989*b*), using the main result of (Moschovakis 1989*a*). Roughly, a recursor

$$\mathbf{int}_A = \mathbf{int}_A^{\mathfrak{M}} \tag{8.3}$$

(the **intension of** A in $\mathfrak{M}$) is naturally associated with each structure $\mathfrak{M}$ and term A, in such a way that the domain of $\mathbf{int}_A$ is a product of the domains of the givens of $\mathfrak{M}$ and its data sets, and its transition and value mappings are defined *explicitly and immediately* using application, conditionals and calls.[33]

Proposal IV: Algorithms are definable recursors *Every algorithm relative to given recursors $\mathcal{F}$ is* (modeled faithfully by) *the intension* $\mathbf{int}_A$ *of some* FLR-*term A on the structure $\mathfrak{M}$ with the givens $\mathcal{F}$; and, conversely, each* $\mathbf{int}_A$ *on $\mathfrak{M}$ is an algorithm relative to $\mathcal{F}$.*

The iterative algorithms of a first-order structure $\mathfrak{M}$ are the $\mathfrak{M}$-intensions of FLI-*terms.*[34]

A functional on the data sets is *recursive* or *computable* relative to $\mathcal{F}$ if it is computed by an algorithm of $\mathfrak{M}$, and it is *iteratively computable* if it is computed by an iterative algorithm of $\mathfrak{M}$. These are exactly the functionals definable by FLR- and FLI-terms on $\mathfrak{M}$, in the natural, denotational semantics of FLR.

Notice that, in a trivial sense, every recursor $\alpha : X \rightsquigarrow W$ models an algorithm of some structure, e.g., the structure (X, W, α)! On the other hand, no function or relation on an arbitrary set M is automatically computable by some algorithm, not the equality relation $x = y$ on M, not even the identity function $f(x) = x$. It is usual, of course, to include such simple functions among the givens of a structure, but it is not necessary—and there are examples where it is not natural.

Definition 8.2 *A structure* $\mathfrak{M}$ *is* **implementable in** *a first-order structure* $\mathfrak{M}'$, *written*

$$\mathfrak{M} \leq_{\mathrm{i}} \mathfrak{M}',$$

if every $\mathfrak{M}$*-algorithm is reducible to some iterative algorithm of* $\mathfrak{M}'$.

In standard, computer science terminology, $\mathfrak{M} \leq_{\mathrm{i}} \mathfrak{M}'$ is expressed by saying that *every recursive program in* $\mathfrak{M}$ *can be simulated by a while-program in* $\mathfrak{M}'$. For the integers, $\mathfrak{N} \leq_{\mathrm{i}} \mathfrak{N}$, but it is not generally true of first-order structures that $\mathfrak{M} \leq_{\mathrm{i}} \mathfrak{M}$, and, in fact, there are natural (infinite) examples in which not every $\mathfrak{M}$-recursive function can be computed by a while-program of $\mathfrak{M}$.[35] The standard reductions of recursion to iteration establish $\mathfrak{M} \leq_{\mathrm{i}} \mathfrak{M}^*$, where $\mathfrak{M}$ is first-order and $\mathfrak{M}^*$ is an expansion of $\mathfrak{M}$ by a stack or other, auxiliary data structures.

A serious attempt to defend and support Proposals I – IV requires a detailed examination of several examples and a comparison of the rigorous theory built upon them with the "naive" theory of algorithms, as it has been developed (especially) by complexity theorists, and this I cannot do here. I will confine myself to a final re-examination of the mergesort, and a brief discussion, in the next section, of the infinitary algorithms which arise naturally in this theory.

The mergesort algorithm The natural structure for the mergesort has data sets L and L^*, and the obvious functions for the manipulation of strings for its givens:

$$\mathfrak{L} = (L, L^*, \emptyset, \mathrm{eq}_\emptyset, \mathrm{head}, \mathrm{tail}, \mathrm{append}, \chi_\leq), \tag{8.4}$$

where $\emptyset$ is the empty string (as a nullary function); $\mathrm{eq}_\emptyset : L^* \to \{\mathit{ff}, \mathit{tt}\}$ checks for equality with $\emptyset$,

$$\mathrm{eq}_\emptyset(u) = \begin{cases} \mathit{tt}, & \text{if } u = \emptyset, \\ \mathit{ff}, & \text{otherwise;} \end{cases}$$

$\mathrm{head}(u)$ and $\mathrm{tail}(u)$ are as in §1 (with $\mathrm{head}(\emptyset) = \mathrm{tail}(\emptyset) = \emptyset$ to make them total functions); $\mathrm{append} : L \times L^* \to L^*$ prefixes a sequence with an arbitrary element of L,

$$\mathrm{append}(x, \langle a_1, \ldots, a_{n-1} \rangle) = \langle x, a_1, \ldots, a_{n-1} \rangle,$$

so that, in particular, $\mathrm{append}(x, \emptyset) = \langle x \rangle$; and $\chi_\leq$ is the characteristic function of the given ordering on L. The basic equations (1.2) and (1.1) refer directly to several, additional operations on strings, but these can be defined (computed) from those of $\mathfrak{L}$, e.g.,

$$\mathrm{test}_1(u) = \text{if } \mathrm{eq}_0(\mathrm{tail}(u)) \text{ then } \mathit{tt} \text{ else } \mathit{ff}$$

tests if $|u| \leq 1$ or not. This is an explicit definition, while $h_1(u)$ and $h_2(u)$ are (easily) defined by recursion, but FLR has a recursive construct and it is quite trivial to write in the end a single FLR term A for the mergesort, with one free variable, a formal version of (6.7) with embedded, recursive terms for the parts $f(u,p,q)$ and $g(v,w,p,q)$. Now the official "model" for the mergesort is the recursor

$$\mathbf{msort} = \mathbf{int}_A : L^* \rightsquigarrow L^*\,,$$

which is assigned to this term by the algorithmic semantics of FLR, and it codes not only how the mergesort depends on the ordering, but the whole "flow of computation" determined by the equations (1.1), (1.2). By the general theory,

$$\mathbf{msort}(u) = f_0(u,\vec{p}) \text{ where } \{p_1(\vec{z}_1) = f_1(\vec{z}_1,\vec{p}),\dots,p_n(\vec{z}_n) = f_n(\vec{z}_n,\vec{p})\}, \quad (8.5)$$

where each of the functionals f_i is defined *explicitly and immediately* from the givens of $\mathfrak{L}$. I have not repeated this full definition here, but in this case[36] it means that each f_i is an application with nesting no more than one-deep, for example

$$f_i(z_1,z_2) = p_j(p_k(z_1),z_2,z_1);$$

or an immediate conditional, for example

$$f_i(z_1,z_2) = \text{if } p_i(z_1) \text{ then } p_j(z_1,z_1,z_2) \text{ else } p_k(z_2);$$

or a truth value $f_i(z_1,z_2) = \mathit{ff}$; or, finally, a direct call to the givens, for example

$$f_i(z_1,z_2) = \mathrm{append}(z_1,z_2).$$

Because the critical given $\chi_{\leq}$ occurs only once in the equations (1.1), (1.2) (when we write them carefully, using the conditional), only one of the f_i's depends on it; and from this it follows that

$$\mathbf{msort}(u) = \alpha(u,c) \text{ where } \{c(s,t) = \chi_{\leq}(s,t)\}, \quad (8.6)$$

where c varies over the set of partial relations $(L \times L \rightharpoonup \{\mathit{ff},\mathit{tt}\})$; $\alpha(u,c)$ is an algorithm of the reduct of $\mathfrak{L}$ which does not include the ordering $\chi_{\leq}$; and (most significantly) *the* where *notation has been extended to make sense when it is applied to arbitrary recursors*, not just functionals.[37] Notice that (8.6) is a *recursor identity*; and once we have it, it is natural to define the *dependence* of $\mathbf{msort}(u)$ on $\chi_{\leq}$ in terms of the dependence of $\alpha(u,c)$ on the partial relation c, from which point on the proof of the basic property of the mergesort follows by the arguments of §1, as we made them precise in §5.

An important aspect of this "finished" analysis of the mergesort is that the application of the method of *parameter variation*, which (maybe) seemed a bit *ad hoc* in §5, arises now naturally from the move from (8.5) to (8.6), one of the basic, general transformations of the theory.

9 Infinitary algorithms

It is clear, from the discussion so far, that, in this approach, there is no absolute notion of *effectively implementable algorithm*, just as there is no absolute notion of *algorithm*, independent of any given. We can only talk of *implementing an* $\mathfrak{M}$*-algorithm* $\mathbf{int}_A$ *in* $\mathfrak{M}'$, meaning that we can find an implementation of $\mathbf{int}_A$ among the iterative algorithms of $\mathfrak{M}'$. At the same time, the theory makes room for the infinitary algorithms discussed in §6, those with closure ordinal greater than ω, which cannot be implemented in any structure whatsoever: what are we to make of them, do they serve any useful purpose, do they help illuminate our intuitive notion of *algorithm*? I will consider here, briefly, two ways in which infinitary algorithms arise naturally, as generalizations of concretely implementable algorithms and as interesting mathematical objects of study, in their own right.

Algorithms on finite structures If we read (6.10) as the definition of an FLR-term C on the expansion $(G, R, =, \exists_G)$ of an arbitrary graph (G, R) by $=$ and the existential quantifier, it yields a nullary algorithm of this structure

$$\mathbf{conn}_d = \mathbf{int}_C : \mathbf{I} \rightsquigarrow \{ f\!f, t\!t\}, \tag{9.1}$$

which, like **conn** of (6.10), computes the connectivity of G,

$$\overline{\mathbf{conn}_d} \Leftrightarrow t\!t \iff G \text{ is connected.}$$

Now $\mathbf{conn}_d$ is somewhat more "detailed" than **conn** (because it also accounts for the explicit steps in the computation), but still, there is some number m such that

$$||\mathbf{conn}_d|| = \text{diameter of } G + m; \tag{9.2}$$

and so, again, $\mathbf{conn}_d$ is implementable if and only if G has finite diameter, by Proposition 7.2. For finite G, we can easily build real, practical implementations of $\mathbf{conn}_d$ which can be programmed on a physical processor—even the trivial implementation suggested in Note 31 is not too bad in this case. These implementations are useful in database theory, but the fact of G's finiteness is hardly used in their construction—typically, we only appeal to it at the very last moment, to build up an implementation of the quantifiers. So, would it help to build a conceptual wall between the implementable and the infinitary definable recursors, fixing the terminology so that the latter would be denied the honorable title of *algorithm*? I would argue that the crucial fact about $\mathbf{conn}_d$ is (9.2) and its Corollary (by (7.2)) that *every implementation of* $\mathbf{conn}_d$ *has closure ordinal* $\geq$ *diameter of* $G + m$, and this has nothing to do with the cardinality of G.

The same considerations apply to much of the work in *structural complexity*, a flourishing area of research in theoretical computer science. It is traditional in this field to study only finite structures, but its basic questions are about algorithms; they often make perfectly good sense on infinite structures as well; and, it seems to me, the field might gain much in clarity (and perhaps even some interesting mathematical results) if people seek answers, at least initially, on arbitrary structures, and put off imposing finiteness conditions until they need them—typically not until the very end of the argument.

The Gentzen algorithm In one of the most celebrated and seminal results of modern logic, Gentzen (1934–35) showed that every proof of predicate logic can be transformed into a *cut-free* proof of the same conclusion, in a canonical proof system based on a few intuitive *natural deduction rules.* The Gentzen *cut elimination operation* is defined recursively, by an equation not unlike that of the mergesort:

$$\begin{aligned}\gamma(d) = {} & \text{if } T_1(d) \text{ then } f_1(d)\\ & \text{else if } T_2(d) \text{ then } f_2(\gamma(\tau(d)))\\ & \text{else } f_3(\gamma(\sigma_1(d)), \gamma(\sigma_2(d))),\end{aligned}$$

where d varies over the set Π of (formal) proofs. The conditions T_1, T_2 and the transformations $f_1 - f_3$, τ, and σ_1, σ_2 are complex, but (at least, in principle) they can be defined explicitly in a natural first-order structure $\mathfrak{G}$ with data sets for formulas, variables, proofs, etc., and the usual syntactic operations on these objects as givens, so that the construction yields a $\mathfrak{G}$-algorithm $\boldsymbol{g} : \Pi \rightsquigarrow \Pi$ with $\overline{\boldsymbol{g}} = \gamma$. An implementation of $\boldsymbol{g}$ (or the similar algorithm invented by Herbrand) is one of the basic routines of every theorem prover.

For classical proof theory, the most important fact about the Gentzen algorithm is that it yields a (cut-free) value $\gamma(d)$ of the conclusion of every proof d, which, together with the special properties of cut-free proofs has numerous metamathematical consequences. Not far behind it is the upper bound of the necessary blowup in the *size* of proofs:

$$|\gamma(d)| \le e(|d|, |d|), \tag{9.3}$$

where the size $|d|$ of a proof is (say) the number of symbols in it and $e(n, k)$ is the iterated exponential,[39] defined by the recursion

$$e(0, k) = k, \quad e(n + 1, k) = 2^{e(n,k)}.$$

Some time after (1934/35), Gentzen (1943) extended this theorem to Peano arithmetic, where, however, matters are considerably more complex because the Gödel Incompleteness Theorem rules out the possibility of a finitary cut elimination result. In a modern version of this construction, we introduce an infinitary version of the Gentzen proof system for arithmetic, whose set Π^* of "formal" proofs includes now some infinite objects and admits an infinitary operation, the so-called *ω-rule*: (roughly) *from the infinitely many premises* $\phi(\mathrm{n})$, one for each numeral n naming a number n, *infer* $(\forall x)A(x)$. The extended Gentzen operation is defined again on Π^* very much like γ, by a recursive equation of the form

$$\begin{aligned}\gamma^*(d) = {} & \text{if } T_1(d) \text{ then } f_1(d)\\ & \text{else if } T_2(d) \text{ then } f_2(\gamma^*(\tau(d)))\\ & \text{else if } T_3(d) \text{ then } f_3(\gamma^*(\sigma_1(d)), \gamma^*(\sigma_2(d)))\end{aligned}$$

$$\text{else } f_4(\lambda(n)\gamma^*(\rho(n,d))),$$

where f_4 is a functional embodying the ω-rule; and this equation, as before, defines a $\mathfrak{G}^*$-algorithm $\boldsymbol{g}^*$, where $\mathfrak{G}^*$ is very much like $\mathfrak{G}$, except that it has a functional embodying the ω-rule. Proofs now have infinite, ordinal length but—and this is the important fact—*the upper bound estimate* (9.3) *persists*, with the extended, iterated exponential on the ordinals; and so Gentzen shows that *every theorem of Peano arithmetic admits a cut-free, infinitary proof of size no more than*

$$\varepsilon_0 = \text{the least ordinal closed under } + \text{ and } \alpha \mapsto \omega^\alpha.$$

There is a large number of metamathematical consequences and by-products of the proof of this fundamental theorem, which rivals the basic Cut Elimination Theorem for its importance.

Now, much of this can be done without ever mentioning the word "algorithm", by dealing directly with the defining, recursive equations, much as we did in §5. But it is not a natural thing to do, and the literature on Gentzen's theorems is full of references to "computational procedures", "constructions" and, in fact, "algorithms". It seems to me that the finitary, implementable, $\boldsymbol{g}$ and its infinitary extension $\boldsymbol{g}^*$ share so many common properties, that it is natural and profitable to study the two of them together, within one, unified theory which takes *recursor structure* and *effective definability* rather than *implementability* as the key, characteristic features of "algorithms".

Notes

1. My first publication on this problem was (Moschovakis 1984), a broad, discursive paper, with many claims, some discussion and no proofs. This was followed by the technical papers (Moschovakis 1989*a*, 1989*b*, 1994*b*) and also (Moschovakis 1991, 1995, 1997), (Moschovakis and Whitney 1995), (Hurkens *et al.* 1998), on the related problems of the logic of recursion and the theory of concurrent processes. My main purposes here are: (a) to return to the original, foundational concerns which motivated (Moschovakis 1984) and re-consider them in the light of the technical progress which has been achieved since then; and (b) to propose (in §7) a modelling of the connection between an algorithm and its implementations, which, in some sense, completes the foundational frame of this program. I have tried hard to make this chapter largely independent of the earlier technical work and as broadly accessible as possible, but I have, almost certainly, failed.

2. I am cheating just a bit here: this re-write is easy if the language can deal with *strings* (as `Lisp` and some extensions of the others do), but a bit cumbersome if we must first "teach" the language the basic operations on strings.

3. A triple $(M, 0, S)$ is a Peano system if M is a set; $0 \in M$; $S : M \to M \setminus \{0\}$ is an injection; and every subset X of M which contains 0 and is closed under S exhausts M. All foundations of the natural numbers start with the facts that: (a) *there exists a Peano system*; and (b) *any two Peano systems are isomorphic*, and differ only in what they do with them.

4. Any two countable, dense linear orderings with no endpoints are order isomorphic (Cantor).

5. In fact, I believe that most mathematical theories (and all the non-trivial ones) can be clarified considerably by having their basic notions *modeled faithfully in set theory*; that for many of them, a (proper) set-theoretic foundation is not only useful but necessary—in the sense that their basic notions cannot be satisfactorily explicated without reference to fundamentally set-theoretic notions; and that set-theoretic foundations of mathematical theories can be formulated so that they are compatible with a large variety of views about truth in mathematics and the nature of mathematical objects. Without explaining it in detail or defending it, the textbook (Moschovakis 1994*a*) applies this view consistently to the presentation of the elementary theory of sets and its applications. The brief remarks here are only meant to clarify what I aim to do with algorithms in the more technical sections of the chapter, following this one.

6. Sometimes we can do more and choose C so that $\sim_C$ is the identity relation on C, notably in the case of Cantor's *ordinal numbers* where the class of von Neumann ordinals has this property. In other cases this is not possible. For example, Cantor dealt with *linear order types* exactly as he dealt with ordinal numbers, but (apparently) there is no way to define in Zermelo–Fraenkel set theory a class of linearly ordered sets which contains exactly one representative from each order isomorphism class. Because of this, "linear order types" can be "defined in set theory" only in the minimal way described here, but their study does not appear to have suffered because of this defect.

7. Not all who adopt it will approve of my description of this view. In his classic (Knuth 1973), for example, Knuth dubs "algorithms" (essentially) what I call "implementations" and avoids altogether the second word. It amounts to the same thing.

8. Still more radical would be to simply *define* "algorithm" to be *constructive proof of an assertion of the form* (3.1), but I cannot recall seeing this view explained or defended.

9. Here "formally" means "without regard to meaning" and not (necessarily) "in a formal language". A coherent axiomatization in ordinary language can always be "formalized," in the trivial sense of making precise the syntax of the relevant fragment of English and the logic; whether (and *how*) the formal version corresponds to the naive one is hard to talk about, and involves precisely the philosophical issues about axiomatizations which I am trying to avoid.

10. Zermelo's axiomatization of set theory is a good example of this. It was first proposed in (Zermelo 1908) quite formally, as an expedient for avoiding inconsistency, and only much later in (Zermelo 1930) was it justified on the basis of a realistic, intuitive understanding of the cumulative hierarchy of sets. By the time this happened, the axioms (augmented with replacement) had been well-established and there was no doubt of their value both in developing (technically) and in understanding the theory of sets.

11. This and the fundamentally set-theoretic nature of (b) in Note 3 are part of the argument for the "necessity" of set-theoretic foundations alluded to in Note 5.

12. It has also been suggested that we do not need algorithms, only the equivalence relation which holds between two programs P and Q (perhaps in different programming languages) when they (intuitively) *express the same algorithm.* It is difficult to see how we can do this for all programming languages (current and still to be invented) without a separate notion of algorithm; and, in any case, if we have a good notion of "program equivalence", we can then "define" algorithms to be the equivalence classes of this equivalence and solve the basic problem.

13. Turing machines are modeled in set theory by finite sets of tuples of some form, but their specific representation does not concern us here.

14. A *partial function* $f : X \rightharpoonup W$ is an (ordinary, total) function

$$f : D_f \to W,$$

from some subset $D_f \subseteq X$ of X into W; or (equivalently) a (total) function $f : X \to W \cup \{\bot\}$, where $\bot \notin W$ is some fixed object "objectifying" the "undefined," so that "$f(x)$ is undefined" is the same as "$f(x) = \bot$". For most of what we do here it does not matter, but the official choice for this paper is the second one, so that "$f : X \rightharpoonup W$" is synonymous with "$f : X \to W \cup \{\bot\}$".

15. The binary representation of a natural number n is the unique sequence $a_k a_{k-1} \cdots a_0$ of 0s and 1s (with $a_k = 1$, unless $n = 0$), such that

$$n = a_0 + 2a_1 + 2^2 a_2 + \cdots + 2^k a_k .$$

16. (Knuth 1973) (essentially, in the present terminology) defines an *algorithm* to be an iterator $\phi : X \rightsquigarrow W$, which also satisfies the additional hypothesis that *for every $x \in X$, the computation terminates.* This termination restriction is reminiscent of the view (IIb) in §3, and it is hard to understand in the context of Knuth's own (informal) use of the notion. Suppose, for example, that Professor Prewiles had proposed in 1990 a precise, mechanical procedure which searched (in a most original and efficient way) for a minimal counterexample to Fermat's last theorem; would we not have called this an "algorithm," just because Prewiles could not produce a proof of termination? And what would be the "meaning" of the Pascal program produced by Prewiles, which (by general agreement) implemented his procedure? It seems more natural to say that Prewiles had, indeed, defined an algorithm, and to say this even now, when we know that the execution of his program is doomed to diverge.

17. In the simplest of the classical, "sequential" methods for implementing recursion, the most important part of the state is a "stack", a finite sequence of pieces of information which (roughly) reminds the machine what it was doing before starting on the "recursive call" just completed. There are also "parallel"

implementations, in which the "stack" is replaced by a "tree" (or other structure) of "processes" which "communicate" among themselves in predetermined ways. This listing of buzzwords is as far as I can go here in suggesting to the knowledgable reader the reduction procedures to which I am alluding.

18. A poset is a structure $(D, \leq_D)$, where the binary relation $\leq_D$ on D satisfies the conditions: (a) $d \leq_D d$; (b) $d_1 \leq_D d_2 \,\&\, d_2 \leq_D d_3 \Longrightarrow d_1 \leq_D d_3$; and (c) $[d_1 \leq_D d_2 \,\&\, d_2 \leq_D d_1] \Longrightarrow d_1 = d_2$. Every set X is a *discrete* poset with the identity relation, $x_1 \leq_X x_2 \iff x_1 = x_2$; and for every W and $\perp \notin W$, the set $W \cup \{\perp\}$ is a *flat* poset, with $x \leq y \iff x = \perp \vee x = y$. Since the empty set is (trivially) a chain and its least upper bound (when it exists) is easily the least element of D, every inductive poset has a least element $\perp_D = \sup \emptyset$. It can be shown that a poset $(D, \leq_D)$ is inductive exactly when it has a least element and every non-empty, directed subset of D has a supremum. There is a tendency in recent computer science literature to widen the notion by omitting the requirement that D has a least element, which is why I am avoiding the common term *dcpo* for these structures. Computer scientists also tend to study only *continuous* (in the appropriate *Scott topology*), rather than the more general *monotone* mappings. This makes the theory easier but it is not general enough to cover all the applications that we need here. The basic facts about inductive sets and monotone mappings can be found in most textbooks on *denotational semantics* and in some set theory books, e.g., (Moschovakis 1994*a*).

19. Various versions of this basic fact have been attributed to different mathematicians, but a special case (with a proof which suffices for the full result) is already a subroutine of Zermelo's first proof of the Wellordering Theorem in (Zermelo 1904).

20. The present version yields, in particular, a natural and comprehensible formulation of *recursor isomorphism*, a notion whose original definition (in (Moschovakis 1989*b*)) is quite opaque.

21. This means that $d_1 \leq d_2 \Longrightarrow \text{value}(x, d_1) \leq \text{value}(x, d_2)$, or, equivalently,

$$d_1 \leq d_2 \,\&\, \text{value}(x, d_1) \text{ is defined } \Longrightarrow \text{value}(x, d_1) = \text{value}(x, d_2).$$

See Notes 14 and 18 for the precise conventions about partial functions.

22. If $t(u)$ is an expression which takes values in $W \cup \{\perp\}$ and in which the variable u occurs, ranging over U, then $\lambda(u)t(u)$ stands for the partial function p, where $p(u) = t(u)$.

23. An isomorphism between two iterators $\phi_i = (\text{input}_i, S_i, \sigma_i, T_i, \text{output}_i)$ $(i = 1, 2)$ on X to W is a bijection $\rho : S_1 \rightarrowtail S_2$ between the sets of states, such that $\rho[T_1] = T_2$; $\rho(\text{input}_1(x)) = \text{input}_2(x)$; $\rho(\sigma_1(s)) = \sigma_2(\rho(s))$, for every $s \in S_1$; and $\text{output}_1(s) = \text{output}_2(\rho(s))$, for every $s \in T_1$ which is *input-accessible*, i.e., such that for some $x \in X$ and some n, $s = \sigma_1^n(\text{input}_1(x))$. The precise result is that *the recursors* $\mathbf{r}_1$ *and* $\mathbf{r}_2$ *associated with two iterators* ϕ_1 *and* ϕ_2 *are isomorphic if and only if* ϕ_1 *and* ϕ_2 *are isomorphic iterators.*

24. This is only approximate; see 'The mergesort algorithm', in §8. Note, also, that we might equally well have set

$$\text{mergesort}_2(u) = p(u) \text{ where } \{q(v,w) = g(v,w,p,q), p(u) = f(u,p,q)\},$$

but mergesort$_1$ and mergesort$_2$ are isomorphic: It is an easy, general fact, that reordering the listing of the **parts** within the braces of a where expression produces an isomorphic recursor.

25. By the simple result quoted in Note 24, however, changing the order in which we specify computations which are to be executed in parallel "preserves the algorithm".

26. Well, maybe not so clear; see the remarks following Proposal IV.

27. Here **I** is some fixed set with a single element (say $\emptyset$), so that a recursor $\alpha : \mathbf{I} \rightsquigarrow \{\mathit{ff}, \mathit{tt}\}$ has no real arguments, and simply computes an object

$$\overline{\alpha} = \overline{\alpha}(\emptyset) \in \{\mathit{ff}, \mathit{tt}, \perp\}\,.$$

I am also using "$\Leftrightarrow$" for the equality relation on $\{\perp, \mathit{ff}, \mathit{tt}\}$ in the definition of **conn**, since $\overline{\mathbf{conn}}$ is a partial relation.

28. This is an outline of the standard proof of Theorem 5.1.

29. In our world, the law is vague and still not fully formed, but (as I understand it) it denies patents to algorithms, but grants copyrights to programs.

30. It is not always true for a monotone τ that $d \leq \tau_x(d)$, but τ_x is only applied to such ds in the construction (6.11) of the iteration sequence

$$\{d^\xi(x) \mid \xi < ||\alpha||\}\,,$$

so that, where it is relevant, applying the transition mapping, indeed, does not decrease our information about the value.

31. You cook up an iterator whose computation for each x is the sequence of iterates $d^0(x), d^1(x), \ldots$, and then check that α is reducible to it. It is, in general, an inefficient implementation, but it is used routinely in finite model theory, and (I have been told) it is also used for some very special database applications.

32. The only non-obvious side condition in the syntax is that in the recursion construct for FLI, the "recursion variable" p occurs only where it is explicitly shown.

33. The language FLR "evolved" somewhat between (Moschovakis 1989*a*) and (Moschovakis 1997), and the intensional semantics are constructed in (Moschovakis 1989) for a more restricted class of recursors, but none of this is very important or affects the present discussion. The mapping $A \mapsto \mathbf{int}_A$ is defined (basically) by recursion on the structure of A, as one might expect. It is not a difficult construction, but it does involve some subtleties and technicalities (mostly in making precise this "explicitly and immediately") which make it impractical to give a useful summary of it here. In addition to the papers already cited, (Moschovakis 1994*b*) discusses some applications of intensional semantics to the

philosophy of language and establishes the *decidability of algorithm identity on any fixed structure with finitely many givens.*

34. The notion of an *iterative algorithm* does not have a clear meaning, except on first order structures.

35. There is a large literature on the reduction of recursion to iteration under various conditions; see, for example, (Tiuryn 1989) and the papers cited there.

36. In fact, the forms listed describe fully the *normal form* for intensions in first-order structures.

37. This follows from the basic, general facts of the theory of recursors, a natural (and not very difficult) extension of the fixed-point theory of monotone mappings which keeps track (in effect) of the recursive definitions, not just their fixed points.

38. The conclusion of the given proof cannot have free and bound occurrences of the same variable, but such details are of no consequence for the point I want to make and I will steer clear of precise definitions and sharp, optimal statements of facts. A good reference for this discussion is (Schwichtenberg 1977), which explains clearly all the results I will allude to—and much more.

39. The precise result is much better than this; see (Schwichtenberg 1977).

Bibliography

Benacerraf, P. (1965). What numbers cannot be. *Philosophical Review*, **74**, 47–73. Reprinted in *Philosophy of mathematics: selected readings* (2nd edn), (ed. P. Benacerraf and H. Putnam), pp. 272–94. Cambridge University Press, 1983.

Frege, G. (1952). On sense and denotation. In *Translations from the philosophical writings of Gottlob Frege* (ed. P. Geach and M. Black). Blackwells, Oxford.

Gentzen, G. (1934–35). Untersuchungen über das logisches Schliessen. *Mathematische Zeitschrift*, **39**, 176–210, 405–31.

Gentzen, G. (1943). Beweisbarkeit und Undbeweisbarkeit von anfangsfällen des transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, **119**, 140–61.

Hurkens, A.J.C., McArthur, M., Moschovakis, Y.N., Moss, L., and Whitney, G.T. (1998). The logic of recursive equations. (To appear in the *Journal of Symbolic Logic.*)

Kleene, S.C. (1952). *An introduction to metamathematics.* North-Holland, Amsterdam and von Nostrand, Princeton.

Knuth, D.E. (1973). *The art of computer programming. Fundamental algorithms*, Vol. 1, (2nd edn). Addison-Wesley, Reading, Mass.

Maass, W. (1985). Combinatorial lower bound arguments for deterministic and nondeterministic Turing machines. *Trans. American Math. Soc.*, **292**, 675–93.

Moschovakis, Y.N. (1984). Abstract recursion as a foundation of the theory of algorithms. In *Computation and proof theory* (ed. M. M. Richter *et al.*), Lecture Notes in Mathematics, Vol. 1104, pp. 289–364. Springer-Verlag, Berlin.

Moschovakis, Y.N. (1989*a*). The formal language of recursion. *Journal of Symbolic Logic*, **54**, 1216–52.

Moschovakis, Y.N. (1989*b*). A mathematical modelling of pure, recursive algorithms. In *Logic at Botik '89* (ed. A.R. Meyer and M.A. Taitslin). Lecture Notes in Computer Science, Vol. 363, pp. 208–29. Springer-Verlag, Berlin.

Moschovakis, Y.N. (1991). A model of concurrency with fair merge and full recursion. *Information and Computation*, **93**, 114–71.

Moschovakis, Y.N. (1994*a*). *Notes on set theory.* Undergraduate Texts in Mathematics. Springer-Verlag, New York.

Moschovakis, Y.N. (1994*b*). Sense and denotation as algorithm and value. In *Logic Colloquium '90,* Vol. 2 (ed. J. Oikkonen and J. Väänänen). Lecture Notes in Logic, Vol. 2, pp. 210–49. Springer-Verlag, Berlin.

Moschovakis, Y.N. (1995). Computable concurrent processes. *Theoretical Computer Science*, **139**, 243–73.

Moschovakis, Y.N. (1997). The logic of functional recursion. In *Logic and scientific method* (ed. M.L. Dalla Chiara *et al.*), pp. 179–207. Kluwer Academic Publishers, The Netherlands.

Moschovakis, Y.N. and Whitney, G.T. (1995). Powerdomains, powerstructures and fairness. In *Computer science logic* (ed. L. Pacholski and J. Tiuryn), Lecture Notes in Computer Science, Vol. 933, pp. 382–96. Springer-Verlag, Berlin.

Schwichtenberg, H. (1977). Proof theory: some applications of cut-elimination. In *Handbook of mathematical logic* (ed. J. Barwise), pp. 867–95. North-Holland, Amsterdam.

Tiuryn, J. (1989). A simplified proof of ddl < dl. *Information and Computation*, **81**, 1–12.

Turing, A.M. (1936). On computable numbers with an application to the Entscheidungsproblem. *Proc. London Math. Soc.* (3), **42**, 230–65. A correction, **43**, 544–6.

Zermelo, E. (1904). Proof that every set can be well-ordered. In *From Frege to Gödel* (ed. J. van Heijenoort), pp. 139–41. Harvard University Press, Cambridge, Massachusetts, 1967.

Zermelo, E. (1908). Untersuchungen über die Grundlagen der Megenlehre I. *Mathematische Annalen* **59**, 261–81. Reprinted as: Investigations in the foundations of set theory I. In *From Frege to Gödel* (ed. J. van Heijenoort), pp. 199–215. Harvard University Press, Cambridge, Massachusetts, 1967.

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae*, **16**, 29–47.

Department of Mathematics
University of California
Los Angeles
CA 90095-1555
U.S.A.

Department of Mathematics
University of Athens
Athens, Greece

email : ynm@math.ucla.edu

5

Truth and knowability: on the principles *C* and *K* of Michael Dummett

Per Martin-Löf

Truth is a highly ambiguous term. At least four clearly recognizable senses, all of relevance for this meeting, can be listed, namely,

> Tarski's notion of truth of a closed formula, or sentence,
> truth of a proposition,
> truth of an assertion, or judgement,
> truth in the sense of reality, as opposed to appearance.

There is an ambiguity in the term 'assertion': you may use it either generally for a claim, a knowledge claim, or specifically for an affirmation, that is, for a claim of the form 'A is true', where A is a proposition. In this talk, I shall use it consistently in the first of these two senses. Also, I shall use the terms 'assertion' and 'judgement' synonymously. At the bottom of my list is the notion of truth as one pole of the distinction between appearance and reality: it is so to speak the high notion of truth, often capitalized, although I have put it at the bottom. At the top is the notion of truth of closed formulas, or sentences, which is the lowest notion in the sense that it is a purely mathematical notion, determined by Tarski's well-known recursive definition, and I shall not be concerned with that either, but remain in the middle region, dealing exclusively with the notion of truth of a proposition and the notion of truth of an assertion, or judgement.

Now, in his long paper 'What is a theory of meaning? (II)' from 1976, Michael Dummett posed the problem, quoting verbatim, of how the notion of truth, within a theory of meaning in terms of verification, should be explained. The idea is of course that, in a truth-conditional theory of meaning, the notion of truth has to be there from the very beginning, since the meaning explanations of the various logical constants are given in terms of truth conditions. But suppose now that we replace the notion of truth as the basic notion by the notion of proof, or verification. Then, at the most basic level, we shall not speak about truth any longer, but instead about proof, or verification, and there then arises the problem that Dummett formulated: even if the notion of truth of a proposition is no longer the basic notion, we are still interested in how it is to be understood. And, in that same paper, he formulated two principles that ought to be satisfied by the notion of truth of a proposition, or statement, as he himself says, namely,

> C: If a statement is true, there must be something in virtue of which it is true,

and

> K: If a statement is true, it must be in principle possible to know that it is true.

Actually, these two principles form a recurring theme in Dummett's writings. The first principle occurs already in his very early paper 'Truth' from 1959, where the formulation is even more explicit, saying as it does that a statement is true only if there is something in the world in virtue of which it is true. Both principles occur together for the first time in the Postscript that was added to it in 1972, and they then recur in Chapter 13: Can Truth be Defined? of *Frege: Philosophy of Language* as well as in 'What is a theory of meaning? (II)', which is where they are labelled C and K.

It is clear from the label of the first principle, C for Correspondence, that it is meant to be a formulation of the well-known correspondence principle, which as we know goes back to Aristotle and is so basic that it so to speak has got to be right if it is sensibly interpreted. For instance, if I say, 'My fountain pen is blue', there is something in the world in virtue of which that is so, if it really is so, namely, the blueness of my fountain pen. So this is it, in this case, which is there and verifies, or makes true, the proposition that my fountain pen is blue. It is clear that the correspondence principle, understood in this very general and unsophisticated way, is somehow right, and has to be right on any conception, whether it be in terms of a truth-conditional or a verificationist, a classical or an intuitionist theory of meaning.

Then there is the second principle K, where I suppose K stands for Knowability, or at least something having to do with Know, which says that, if a proposition is true, it must be in principle possible to know that it is true. As you see, this is a principle which is quite different from C, and, whereas C is so to speak readily accepted, when you look at K, I think you immediately get some feeling of uneasiness: could the 'true' that you have in the conditional clause possibly be the same 'true' as you have in the main clause? It sounds somehow strange to say that, if a proposition is true, then, from that alone, it follows that it is in principle possible to know that it is true, in the same sense of 'true' as you have in the conditional clause: it seems somehow unlikely. At least, this has left me with uneasiness, and the purpose of this talk is to try to resolve the difficulties which are inherent in the principle K, and actually also to emend it in such a way that it becomes acceptable.

The key to resolving the perplexities surrounding the principle K turns out to be the very distinction between the notion of truth of a proposition and the notion of truth of an assertion, or judgement, with which I started. First of all, I should say that you cannot hope to explain these two notions, truth of a proposition and truth of a judgement, in isolation: they are two concepts that fit into a certain conceptual structure, where also other notions are involved, and, if we want to clarify them, we shall have to display, as it were, this little conceptual structure, or conceptual system, and see how the various pieces fit together and what functions they fulfil in it. The key elements of this conceptual structure are the ones that are displayed in the following table:

Non-epistemic concepts	Epistemic concepts
proposition	judgement
proof (verification) of a proposition	proof (demonstration) of a judgement
truth of a proposition	truth (correctness) of a judgement

So we shall have to clarify the notion of proposition and the notion of judgement, and we shall have to clarify the notion of proof of a proposition as opposed to the notion of proof of a judgement. Here we have a good terminological possibility, because in English we have both the term 'proof' and the term 'demonstration', and 'demonstration' is quite unambiguously associated with making something evident, which is to say that it is an ideal word to use on the epistemic side, demonstration of a judgement, and then we get the term 'proof' free for propositions, or, if you prefer, you could also use verification in connection with propositions. Finally, we shall have to clarify the two notions with which I started, namely, the notion of truth of a proposition and the notion of truth of a judgement, and, if one finds it inconvenient to use truth in both cases here, although it is sometimes unavoidable, one can decide to use correctness, or objective correctness, in connection with assertions, or judgements: that is what Dummett usually does. But, of course, this means already deciding to make a technical distinction between truth and correctness, because no doubt *ἀληθής* and *ὀϱθός* were used essentially synonymously in connection with the Greek *δόξα* and, similarly, in scholastic philosophy, you had the Latin *judicium verum seu rectum.* But now that we have the two words, this is a convenient technical terminology.

To begin with, I would like to say something preliminary about the distinction between propositions and judgements, before properly answering the questions, 'What is a proposition?' and, 'What is a judgement?' Now propositions are the things that are held true, or sometimes held false, and the things on which the logical operations operate: the connectives operate on propositions and the quantifiers on propositional functions. Judgements, on the other hand, are what we demonstrate: in each step of a chain of reasoning, or demonstration, we proceed from some previously demonstrated judgements to a new judgement, which is evident on the grounds of the previous ones, such a step being an inference with the previously made judgements as premises and the newly made judgement as conclusion. The forms of judgement are totally different from the logical operations. First of all, we have the affirmative form of judgement

$$A \text{ is true,}$$

where A is a proposition. This is the only form of judgement that I shall need to consider in the course of this talk, but there are also hypothetical judgements, or consequences, general judgements, and hypothetico-general judgements, which all have their own characteristic forms. A Gentzen sequent is an example of a hypothetico-general judgement, that is, a judgement which is both hypothetical and general. These are some forms of judgement that are used in predicate logic, but there are many more forms of judgement, and, just as we cannot limit in advance our logical operations, we cannot limit in advance our forms of judgement: indeed, in type theory, there are several other forms of judgement, in particular, the form of judgement which is used to say that something is an object of a certain type, and, just as crucially, that two objects of a certain type are the same, where 'same' means definitionally or intensionally the same.

In the preceding, I made the distinction between propositions and judgements in a preliminary fashion by simply giving examples of some well-known forms of proposition and some well-known forms of judgement, but, by doing so, I have of course not really defined what a proposition is and what a judgement is. So what is a judgement? Well, the notion of judgement, and everything actually that stands in the right-hand column of my table, is an epistemic notion, which means that the notion of knowledge is crucially involved. The simplest answer to the question of what a judgement is seems to be to say that a judgement is defined by laying down what it is that you must know in order to have the right to make it. Or, using the term 'assertion' rather than 'judgement', an assertion is a knowledge claim, and hence, in order to clarify the assertion, you have to clarify what knowledge it is that you claim to have when you make the assertion. So, however you phrase the explanation, the crucial question is, 'What knowledge?'

Now, once we have fixed the notion of judgement in this way, the notion of demonstration of a judgement, which is located on the second line of the right-hand column of my table, is defined simply by saying that a demonstration is what makes a judgement known, or evident: a demonstration is a chain of reasoning, and what it purports to do is to make the final judgement of that chain known, or evident. There are many words that you can choose among here, and from a logical point of view it is immaterial which of these terms you choose, because they are but different labels of one and the same piece in the conceptual structure, and you may label that piece in any way you want: the only important thing is how it functions in the structure. The natural labels here are to say known, evident, demonstrated, justified, or warranted: this is the term usually adopted by Dummett, contrasting as he does an assertion's being warranted with its being correct, which is the next notion to be analysed, or you may say reasoned, or grounded. So the notion of evidence here comes before the notion of truth, or correctness, of a judgement in the conceptual order.

But now, having the notion of evidence, or knownness, how do we define the notion of truth, or correctness, of a judgement? Well, the proper conceptual connection seems to be this: a judgement is by definition true, or correct, if it *can* be known, or made evident. You see, evident means known, which is to say, actually known, but a judgement is true, or correct, if it is knowable, evidenceable, demonstrable, justifiable, warrantable, or groundable, whichever you prefer. The crucial notion that comes in here is the notion of possibility, and it is of course a question of possibility in principle. So the difference between, on the one hand, known, evident, demonstrated, and so on, and, on the other hand, knowable, evidenceable, demonstrable, and so on, is nothing but the difference between actuality and potentiality. Now this definition of the notion of truth, or correctness, of a judgement validates Leibniz's principle of sufficient reason. The most widespread formulation of it has several ingredients, but, if we restrict ourselves to what has to do with the truth of judgements, then what Leibniz's principle of sufficient reason says is that, if a judgement is true, then it can be known. A judgement is not true unless there exists a reason for it, that is, unless a reason for it can be given: that is the content of the principle of sufficient reason. And why does it hold? Well, it holds because of the definition of the notion of truth of a judgement: truth of a judgement is simply defined as knowability, and therefore the principle holds. This was also Leibniz's own view, that the principle of sufficient reason is contained in the definition of the notion of truth.

Now, as an indication that the conceptual connections have been properly made here, I would like to say a few words about Descartes' criterion of truth. Stated as briefly as possible, it says that, if a judgement is evident, then it is true: *si quid intellectui meo sit evidens, illud omnino est verum.* There is no surer sign of the truth of a judgement than our having made it evident to ourselves: that is the gist of Descartes' truth criterion. So evidence implies truth, or correctness, of a judgement. Now, as Brian McGuinness said in his introduction to this meeting, Descartes had to invoke the veracity of God in order to justify his truth criterion, because why does it hold, according to Descartes? Well, it holds because he took it as an axiom that God does not deceive us. But, at least to my mind, it would be very strange if one should have to invoke the notions of God and deception in order to see that the evidence of a judgement entails its truth. Things of this sort normally hold on purely conceptual grounds, and you see now how it comes out: truth is simply defined as evidenceability, and hence Descartes' truth criterion, saying as it then does that, if a judgement has been made evident, then it can be made evident, follows from the principle that, if something has been done, then it can be done. This, on the other hand, is the truly fundamental metaphysical principle which was given the succinct scholastic formulation *ab esse ad posse valet consequentia*, a formulation which in its turn probably derives from the short passage *ἔξ ἐνεϱγείας ἥ δύναμις* in Aristotle's

Metaphysics, Book Θ, Chapter IX. So Descartes' truth criterion, fundamental as it may seem to be, is actually a consequence of this even more basic principle, the *ab esse ad posse* principle.

Another effect of this definition of the notion of truth of a judgement is that the traditional Platonic characterization of knowledge as justified true opinion, *δόξα ἀληθὴς μετὰ λόγου,* opinion true with justification, or by aid of justification, does not look natural any longer when the notion of truth receives the conceptual determination that I have just given to it. Indeed, since true is the same as justifiable, justified true opinion becomes justified justifiable opinion. But, if an opinion is justified, it is superfluous to say that it is justifiable by the *ab esse ad posse* principle. Hence 'justifiable' can be omitted from the formulation, and we get the simpler characterization of knowledge as justified opinion. Also, although *δόξα* is traditionally rendered by opinion, it is equally well translated by judgement, so a piece of knowledge is the same as a justified, or demonstrated, judgement. Presumably, the reason for the more complicated formulation is that, from Plato onwards, the notions of knowledge and truth have been associated with infallibility, and, if you include infallibility in the notion of truth of a judgement, then you cannot argue from evidence to truth in this simple way by the *ab esse ad posse* principle, and that is precisely why Descartes had to invoke the veracity of God at this point. Now, as a matter of fact, our demonstrations are not infallible: a demonstration purports to make something evident to us, and it is the best guarantee that we have, but it is not infallible. We do sometimes make mistakes in our demonstrations, and hence, if you include infallibility in the notion of truth of a judgement, then the step from evidence to truth cannot be taken any longer. That means that the problems that have to do with infallibility have to be moved to another level, so to speak, and that is the level that I put at the bottom of my list, that is, the highest level that has to do with the notion of truth in the sense of reality as opposed to falsehood in the sense of appearance, illusion, or deception, and that will be completely left out of my talk.

This finishes the semantical explanations of the concepts occurring in the right-hand column of my table, that is, the epistemic concepts that are associated with the notion of judgement. There remain the non-epistemic concepts in the left-hand column of the table, which is to say, the notion of proposition, the notion of proof, or verification, of a proposition, and the notion of truth of a proposition. So what is a proposition? Well, in a truth-conditional theory of meaning, a proposition is defined by its truth conditions, whereas, in a verificationist theory of meaning, this explanation is replaced by saying that a proposition is defined by its proof conditions, or verification conditions, which state what a proof, or verification, of the proposition looks like. Now it has sometimes been said, for example, by Dummett in his paper 'Truth' from 1959, that the difference between a classical and an intuitionist, or constructivist, account of the meanings of the logical constants is that truth conditions are replaced by assertion conditions. But observe that that is not what I am saying here: I am saying that truth conditions are replaced by proof conditions, or verification conditions. Now the notion of assertion condition is also important, but the role of an assertion condition is to determine the meaning of an assertion, or judgement, the concept that we had at the top of the right-hand column of the table of concepts to be explained: an assertion, or judgement, is defined by its assertion condition, that is, by laying down what it is that you must know in order to have the right to assert it.

As concerns the notion of proof of a proposition, we must distinguish between proofs, or verifications, of the forms that enter into the meaning explanations of the various logical constants on the one hand, and arbitrary proofs, or verifications, on the other. We all know the Brouwer–Heyting–Kolmogorov explanations of the meanings of the logical constants, which run according to the pattern: a proof of a conjunction $A \,\&\, B$ is a pair consisting of a proof of A and a proof of B, and similarly for the other logical operations. But we also have to allow proofs which are not directly of one of the forms that enter into the meaning explanations of the logical constants, just as, when we let the natural numbers be defined by the first two Peano axioms, 'Zero is a natural number', and, 'If n is a natural number, the successor of n is a natural number', some innocent person may come and ask, But what about $2 + 2$, is it not a natural number?' The answer is of course that, when you give an inductive definition, like that of the natural numbers, it is tacitly understood that something should count as a natural number even if you may need to calculate it a few steps to get it into zero or successor form, and similarly here, in the Brouwer–Heyting–Kolmogorov explanations, a proof in general may have to be calculated before you get it into the form, or one of the forms, that define the proposition in question. We then have two terminological possibilities, either to call proofs of the forms that enter into the meaning explanations of the logical constants simply 'proofs', in which case we would in general only have a method of proof, or to call proofs of the forms prescribed by the meaning explanations 'canonical proofs', or 'direct verifications', in which case we also have to allow non-canonical proofs, or indirect verifications. Choosing the latter alternative, a non-canonical proof, or indirect verification, becomes clearly the same as a method of canonical proof, or direct verification. So we have these two terminological possibilities.

Now, if a proposition is defined in this way by its proof conditions, then, when you come to the next question in the left-hand column of my table, which is to say, 'What is a proof of a proposition?', the answer is exceedingly simple, because a proposition was defined precisely by laying down how its proofs are formed, which means that there is nothing more that needs to be said. Indeed, once we have understood the proposition, we already know what a proof of it is, a canonical proof in the first place, and then a proof in general is a method such that, when you execute it, you obtain a canonical proof as result.

There now remains in the left-hand column only the notion of truth of a proposition, which appears on the third and last line. So we must ask ourselves, 'How is the notion of truth of a proposition to be defined?' This is precisely the problem of Dummett's that I started by quoting, namely, of how the notion of truth, within a theory of meaning in terms of verification, should be explained.

The answer is most simply given in the form of the chain of equations

$$\begin{aligned} A \text{ is true} &= \text{there exists a proof of } A \\ &= \text{a proof of } A \text{ can be given} \\ &= A \text{ can be proved} \\ &= A \text{ is provable,} \end{aligned}$$

in which the equality sign signifies sameness of meaning. So here again the notion of possibility in principle comes in, but now it is in connection with the notion of truth of a proposition, whereas previously it was in connection with the notion of truth, or correctness, of a judgement. And 'proof' is here to be understood in the sense of 'canonical proof', which means that the truth of a proposition is equated with the possibility of coming up with a canonical proof, or direct verification, of it. So here I have chosen the first of the two terminological alternatives that I mentioned. Now what are given in the chain of equalities are but different permissible readings of one and the same form of judgement 'A is true'. After all, I have to follow my own official explanations, and, since 'A is true' is a form of assertion, or judgement, its meaning is determined by laying down its assertion condition, that is, by laying down what it is that you must know in order to have the right to make a judgement of this form, and, in this case, the explanation is that, to have the right to make a judgement of the form 'A is true', you must know a proof of A, a proof which is in general non-canonical, that is, which is in general merely a method such that, when you execute it, you get a canonical proof as result. Now that is the official meaning explanation, but it is clear from that meaning explanation that you may allow yourself to read 'A is true' in these different ways, which are of course quite similar, actually, to the reading that Kleene used in his realizability interpretation: forgetting about all other differences, Kleene read the proposition that there exists a realizer of A, where A is an arithmetical formula, as 'A is realizable'.

Now let me return to my original promise of clarifying Dummett's principles C and K. If you first look at C, 'If a statement is true, there must be something in virtue of which it is true', you will see that it is in complete agreement with what I have said about the notion of truth of a proposition. Indeed, the verificationist definition of truth is that a proposition is true if there exists a proof of it, so, if we just call that something in virtue of which a statement is true its proof, or verification, then C is nothing but the definition of truth that I just gave. That means of course that the intuitionist, or verificationist, notion of truth is really a version of the correspondence notion of truth, truth as agreement with reality: the only novelty is that we call that thing in reality, or in the world, which has to be there in order for the proposition to be true, its proof, or verification.

Let us now finally turn to the principle K, 'If a statement is true, it must be in principle possible to know that it is true'. So remember the principle of sufficient reason, which says that, if a judgement is true, then it can be known. Now apply the principle of sufficient reason to a judgement of the particular form 'A is true', where A is a proposition. Then what we get is that, if a judgement of the form 'A is true' is correct, then this judgement can be known, but that is the same as saying that the proposition A can be known to be true. So we have now achieved in the main clause exactly what you find in the principle K, but there is a fundamental difference in the conditional clause, which no longer takes the simple form 'if the proposition A is true', but the more complicated form 'if the assertion, or judgement, "A is true" is correct'. This means that two truth operators have turned out to be involved here: one is the truth of the given proposition, and the other is the truth, or correctness, of the judgement which is obtained by applying the first truth operator to the given proposition. So this is the corrected form of the principle K that we have arrived at:

> If a judgement of the form 'A is true' is correct, then the proposition A can be known to be true.

Now, unfortunately, this reads a bit awkwardly, but it may be rephrased in the following way, if only we accept the principle that a judgement of the form 'A is true' is correct if and only if the proposition A really is true. This is a principle that I think everybody accepts: the only difference that you find between the realist and the idealist is in the sense that they give to the qualifier 'really' that appears here. The realist takes that notion as a primitive notion that cannot be reduced to anything else, whereas, on the analysis that I have given, the notion of reality that comes in here is nothing but the notion of knowability. In any case, the principle is acceptable as it stands, and hence we can replace saying that the judgement that A is true, where A is a proposition, is correct by saying that the proposition A really, or in reality, is true. If we make that replacement, we arrive at the following

> Emendation of K: If a proposition really is true, then it can be known to be true.

This is the amended version of the principle K that I propose. It agrees entirely with the principle K in the main clause, but has a crucial modification in the conditional clause, and it is an almost immediate consequence of the principle of sufficient reason.

Acknowledgements

This paper is an only slightly altered transcript of the talk that I gave at the conference, to whose organizers I am indebted for a tape recording which lasted until the power failed. After that, I have relied on a battery-powered tape recording, kindly given to me by Margaret MacDougall.

Bibliography

Dummett, M.A.E. (1959). Truth. *Proceedings of the Aristotelian Society*, new series, **59**, 141–62. Postscript (1972). In *Logic and philosophy for linguists: a book of readings* (ed. J. M. E. Moravcsik), pp. 220–5. Mouton, The Hague.

Dummett, M.A.E. (1973). *Frege: philosophy of language.* Duckworth, London.

Dummett, M.A.E. (1976). What is a theory of meaning? (II). In *Truth and meaning* (ed. G. Evans and J. McDowell), pp. 67–137. Clarendon Press, Oxford.

Department of Mathematics
University of Stockholm
106 91 Stockholm
Sweden

PART II

Formalism and naturalism

6

Logical completeness, truth, and proofs

Gabriele Lolli

When discussing such topics as 'truth in mathematics', it would be culpable negligence to ignore what logic has to say on the subject. What logic has to offer are theorems, not speculations. We teach these theorems in logic classes, but apparently to no avail, since the teaching does not seem to leave any trace in grown-up mathematicians. Perhaps this is due to the fact that when proving theorems their significance is seldom discussed; only their mathematical usefulness, as opposed to general wisdom, is stressed. Theorems are admittedly not a detailed description of reality; they refer to idealized models; logic deals only with models of reasoning; but mathematicians, and scientists in general, should know what a tremendous amount of reliable and useful information is conveyed by properties of abstract models. We should sometimes pause to reflect on theorems, besides proving new ones, especially when, as is the case in logic, they concern our own activity.

One theorem that is included in all introductory courses, perhaps the only one always proved, is the completeness theorem. It deals precisely with questions of truth; we will discuss what it has to teach about mathematical activity and refer to Lolli (1995) for a more general appraisal of its import on semantic matters.

1 Logical truth

The starting point of these reflections is that there is no place for truth *in* mathematics, no so-called *mathematical truth.* If we look at (statements that are labelled) theorems in any written mathematical text, in no one of them do we find the word 'true'. Theorems are statements which are accompanied by arguments—so-called proofs—linking them to others and eventually, by iteration, to statements called axioms of a theory. The link is that of logical consequence. There is no theorem which upon careful inspection does not turn out to be a logical consequence of the axioms. Hence every theorem is a statement B for which there is another statement A which is a conjunction of a finite number of the axioms of the theory and such that $A \to B$ is logically true.

This is a fact, not an *a priori* definition of mathematics, although it could be converted into one. Actually, it has been done, not by formalists but by mathematicians of quite different philosophical bents. At the turn of the century mathematicians had to make sense of the axiomatic method, never before so powerful and pervasive. Scholars with such a variety of outlooks as Poincaré, Hilbert, Enriques, Pasch, Peano, and others all gave this same characterization of theoremhood, even if they did not use our logical terminology. We recall just one example, dating from still unsophisticated times. Moritz Pasch (1882, p. 98) warned that in order for geometry to become a truly deductive science it was necessary that the derivations of consequences be independent of the sense of geometrical concepts, as they were from pictures.

> In the course of a deduction, it is allowable, and it can be useful to think of the meaning of the involved geometrical concepts; but it is by no means necessary; when it becomes necessary, it is a symptom of a defective character of deductions, and of the inadequacy of the propositions assumed for the proof.

He also stressed that, if a theorem is deduced from a set of assumed propositions, called generators, then the value of the deduction surpasses the original goal: by changing the geometrical concepts in the generators, with no supplementary work one gets a new proposition which is a consequence of the transforms of the generators.

It has been argued that the logical definition of theoremhood does not exhaust the philosophy of the founders of modern axiomatic methods, but when you have said what a theorem is, it is difficult to see what else you need.

It could also be argued that the logical definition would be a good one, since it would be consistent with, and would take seriously, the fact that so-called mathematical statements are *formulae.* It is their being formal words devoid of any meaning that allows and explains multiple interpretations; hence the applications of mathematics. There is nothing to be gained by substituting a fixed abstract meaning to the plurality of potential meanings. It only makes it more difficult to come back to the concrete ones in ordinary applications. Nobody knows what it really means to substitute a meaning, unless by passing through an invariant form. In his confused way, Pasch was quite clear about this.

Any other characterization of theoremhood, especially in terms of truth (either objective, say in a universe of sets, or conveyed by assertability conditions, for example, by a truth-founding notion of proof) introduces from the outside unnecessary concepts. When a mathematician says that some proposition is true in the universe of sets, he (or, once for all, she) is stating a theorem of ZF; he could have in mind a model of ZF with additional structure, but if he is capable of describing it then he is doing nothing else than enunciating a new axiom, and its consequences. If not construed in an eliminative way, founding concepts such as truth by Tarski's analysis are bound to import (either vagueness and unregimented intuition or) higher and higher abstract notions in an unending ascent.

Logical truth itself is not an easy notion. But logical truth is *not* a notion of truth and does not require any definition of truth. 'Logically true' means 'true under any interpretation whatsoever', and this implies 'true under any notion of truth'—granted only compositionality with respect to the logical particles.

Logical truth is all kinds of truth, hence no special truth. An interpretation might be a mapping in a set-theoretical structure as it could be a translation in a natural language, each with its own more-or-less explicit notion of truth. Truth is a concern of the applied mathematician, or of the scientist, or of the layman as well, for each particular interpretation. We do not want to get entangled in the question of uniqueness of truth, but it is likely that truth of common-sense physics is quite another truth than that concerning elementary particles.

We are referring of course to logical truth as defined in contemporary mathematical logic. Logical truth has been tailored to the needs of the mathematical method. The multiple-interpretation version comes from the practice and theory of the axiomatic method, and it is by no means a natural one (though one can find some anticipation as far back as Aristotle). In the nineteenth century there were still at least two other competing notions in logic, that of necessity and that of the laws of thought. They were hard to define, and with mathematical explication they faded away; as usual with a scientific definition, something of the intended meaning has been lost, but this does not mean that we cannot still be interested in it; actually it lurks in our minds when we tend to attribute to logic the force of a conclusive argument, or we take it as the banner of rationality.

The great success of mathematical logic is to have shown that all of logic is independent of a definition of truth (and luckily so, since the latter is undefinable). In modern logical theory, truth is a technical but dispensable device, as has been explained by Quine (1970, Chapter 3). To say that a statement is true is tantamount to uttering and asserting the statement. This does not mean that a truth predicate is useless:

> where the truth predicate has its utility is in just those places where, though still concerned with reality, we are impelled by certain technical complications to mention sentences... [and this happens in particular] where we are seeking generality.

For example, 'Tom is mortal or Tom is not mortal' cannot be generalized by quantifying over humans (as with 'Tom is mortal', 'John is mortal',... , and 'all men are mortal'), nor by quantifying over predicates, because the statement is somehow about a sentence and not about reality. We want to say that every sentence of the form 'p or not-p' is true, so we have to quantify over sentences. These become, through quotation, terms in a meta-language, and we need then a device of disquotation, the truth predicate, to get the truth conditions: 'p' is true if and only if p. Unfortunately, by Tarski's analysis, the truth predicate 'is not eliminable by any facile paradigm, only in a devious way if some powerful apparatus is available'. But it is, in a roundabout way: we substitute the statement that every sentence of the form 'p or not-p' is true with the statement that the formula p or not-p is logically true, and we quantify over interpretations instead of over sentences; then to prove equivalence we use the completeness theorem in the substitutional version (a formula is valid if and only if every substitution of its schematic letters with formulae of a language—say arithmetic—transforms it in a true sentence of the language).

2 Proofs

Since we started with questions of fact, the fact cannot be denied that 'truth' has a prominent place in the discourses of mathematicians, if not in their written texts. They use it both in the course of conceiving a theorem and in discussing their progress with colleagues, and when philosophizing. The latter can be dismissed, since it seldom attains the level of a consistent position; see, for example, (Penrose 1994), where the most recurrent words are 'unassailable truths', to refer to properties of the natural numbers, supposedly accessible to the human mind while inaccessible to machines (see also (Putnam 1995)). When philosophizing, mathematicians are caught in the coils of the sterile traditional philosophical schools, as has been denounced by Hersh (1979). We believe that it is possible to make sense of the way that mathematicians talk of truth, while at the same time saving the central place of logic in the definition of mathematics.

When mathematicians assert their statements 'B is a theorem of theory A', or '$A \rightarrow B$ is logically true', they might be telling the truth or not. If there is no truth *in* mathematics, there is a problem of the truth *of* mathematics. While theorems have no truth, the fact that a statement is a theorem is a fact that can be either true or false. Mathematics in polished form is a kind of speech, and we could simply assert '$A \rightarrow B$ is logically true' in the normal flow of discourse. But to add emphasis, we often say 'we assert that $A \rightarrow B$ is logically true', 'it is a fact that $A \rightarrow B$ is logically true', 'it is evident that $A \rightarrow B$ is logically true', or sometimes 'it is *true* that $A \rightarrow B$ is logically true'; it is the italicized 'true' that is in need of explanation. It is not pleonastic, as the logical theory would have it. Or it is, from a logical point of view, but the emphasis suggests that there is more. It warns that we have done something to verify the statement before asserting it. We do not want to confuse truth and verification; verification is needed here as it is for all scientific statements, but verification has a very special character, because of the special character of statements of logical truth.

The truth of '$A \rightarrow B$ is logically true' is not a mathematical truth, it is a truth of this world, though a world enriched with abstract entities (such as linguistic *types*). It says that if we take any interpretation whatsoever, we can safely anticipate that $A \rightarrow B$ will come out true under that interpretation. The assertion certainly does not describe any simple empirical fact; it shares more of the nature of natural laws, at least in the sense of having a universal character. Universal true propositions (like 'all humans are mortal') are never checked directly, say by enumeration, or inductively; they are always established in an indirect way, for example deduced by some other more basic physical laws. We are not allowed this move for '$A \rightarrow B$ is logically true', unless to delay things, since at most we would only possibly find some other hypothesis to add to A, if we had overlooked some. Logical truth on the other hand can seldom be shown to hold by a straightforward verification of the definition, the latter being infinitary and concerning such vague entities as interpretations.

Here comes, as a blessing, the completeness theorem, together with twenty five hundred years of mathematical practice. The theorem asserts that whenever

$A \rightarrow B$ is logically true, there is a proof of it—an amazing fact not true for other kinds of universal laws, or we would have a strong case for omniscience. It has been remarked by Kreisel (1967) that the completeness theorem also shows that well-defined set-theoretical interpretations are sufficient, but this is no decisive gain—if it is not a loss, as far as the richness of interpretations is concerned: logical truth remains infinitary, although better mathematically defined, with respect to set-theoretical interpretations.

In fact mathematicians are usually busy at building proofs, not just uttering sentences; the mathematical discourse is not made only of statements of the form $A \rightarrow B$, but of these accompanied, as we said, by proofs. This is why a theorem is introduced with a warning, that the author is going to establish it: what he means and he says is that 'there is a proof that $A \rightarrow B$ is logically true', and then he always proceeds to give one.

The statement '$A \rightarrow B$ is logically true' must be true in the same sense as any scientific statement (we fully endorse the plea of Maddy (1998*c*) for naturalism): all scientific truths must in a sense be proved; but it is difficult to define 'proof' in general; it is of the same order of difficulty as defining the scientific method. One could amass several specifications, say that a proof is a logical argument, possibly supported by experiments, rationally convincing, giving irrefutable reasons to accept its conclusion. But these notions, having to do both with objectivity and with belief-building, are undefinable. The plain fact is that the prover argues until people say: 'OK, we are convinced'. A logical proof is something different. First of all it is homogeneous to the statement to be proved, being made of the same kind of formulae as the conclusion itself.

A proof is a finite structured object; we can also verify it mechanically. If we are looking for a proof, we can search in a space of finite objects. But thanks to the validity of the rules a proof is a label bearing the notice: 'This is to certify that $A \rightarrow B$ is logically true'. If we recognize and accept the object as a proof, we have as an extra bonus logical truth. This is the good news: logical truth becomes inter-subjectively and effectively checkable when a proof is given. Its overall structure allows direct checking of the relevant syntactic features, instead of infinitely many interpretations. When sufficiently detailed, proofs allow the verification of (the truth of) something being a theorem to become a kind of empirical search and verification in the space of finite combinatorics (though not necessarily confined to combinatory means). They may have other functions, but this one certainly cannot be denied.

So this is the good news, but good news is always accompanied by bad news. Through the above argument, the main theme has shifted from that of truth in mathematics to that of *proof* in mathematics, with its web of questions: formal *vs.* informal proofs; how formal a proof can be; how formal a proof should be; surveyability of proofs; long proofs; mechanical proofs; what kind of confidence a proof can give, what kind of certainty, and so on and so forth. There has been a lot of discussions on these topics in the last years, especially in connection with automation of proofs and computer-assisted proofs; I shall not try to summarize the discussion, but shall concentrate on two points (as a sample of the literature see (Appel and Haken 1986), (Davis and Hersh 1980), (Lolli 1986), (Detlefsen 1992), (MacKenzie 1995), (Swart 1980), and further references therein).

The first bad news is that 'finite' is an elusive notion; it can easily convert itself into 'exceedingly long' and 'unmanageable' in the trade-off between precision and length. One could safely say from this point of view that perfect formal proofs do not exist. If they exist at all, they exist within a computer, be it made of silicon or of neurons, and we cannot put our hands on them. Verification becomes again a matter of indirect persuasion.

Not only do formal proofs not exist in a surveyable sense, but they do not seem to exist even intentionally, since mathematicians do not conceive their proofs as following logical rules. No mathematician, indeed no sane person, makes logical deductions according to codified rules, but for those steps so easy and natural that any person does them also in real life. But what is more damning is that mathematicians do not seem to be aiming at building a logical proof. They do not see themselves as craftsmen building finite objects. Their aim is to convince their peers, to have them repeat their mental processes, in order to lead them to see as they see the inescapable conclusion. Notwithstanding the lip service paid to logic, proof is apparently all that has been left out of the logical definition of consequence and proof.

Attitudes and practices concerning proofs being what they are, one can only rejoice at their success; problems are not to be solved by decree; we cannot impose uniform and unnatural standards. (See Manin's contribution (1998) to this volume for a balanced view.) We can only try to make sense of the situation by ensuring that there is no contradiction in the tension between formal and informal proofs. Formal proofs are the ideal objects asserted to exist by the completeness theorem; by informal proofs we mean those arguments that are explicitly produced by mathematicians.

3 The mathematician as a meta-mathematician

Here again the completeness theorem comes to the rescue. The theorem says that, *no matter* how it is established or accepted, the fact that a sentence is a logical truth, say that God sees it, or a Greek oracle says it, implies that its proof exists.

What people, who are not gods, do is to argue and to reason. That is their nature as language-speaking beings. They communicate through words and pictures and gestures. Since thought has nothing but itself to reach the aim of building a proof, we would expect that what is offered as a proof is just the report of the actual reasoning that has been done. We would expect the actual reasoning to bear some structural resemblance to the final object, which is itself (called) a linguistic text. But it need not be so. The objects need not resemble the machinery they are produced by. Final formal proofs are a kind of limit to which the proofs tend through a tangled web of abstract, infinitary, intensional reasoning, just as nets of sets converge to points.

When a mathematician argues for a theorem he can use whatever type of argument he prefers and considers valid. And he can change style in progress. He can start by noting that something easily follows from the axioms (as when one begins group theory by proving the uniqueness of the neutral element), then pass to the semantic definition and talk of structures and operate on them. Later he can interrupt this line of thought, and start doing computations, thus following formal rules, or even insert a computer printout. According to the audience, he might rely more on geometrical insights or else expound analytical considerations. He can stick to the language of the problem or explore what Hilbert called impure demonstrations. It does not matter. When his actual reasoning is accepted by the community in the large (not only by restricted groups of specialists) it is always the case that this reasoning is recognized as being or containing a sufficiently detailed set of instructions to build a proof.

The existence of the formal proof is guaranteed also if the mathematician did not intend to reach one, if he has for example a purely rhetorical conception of logic. Of course when a proof is accepted the community as a whole could have been led astray by the prestige of the author, or by other factors such as the will to believe, but this is not a mathematical fact, it is a sociological fact.

It is usually said: 'the proposed argument could be converted into a formal proof, given enough time and resources and patience'. But the conversion need not consist only in breaking a step into smaller steps, or in expanding modules. It could contemplate a call to the completeness theorem itself, to replace a whole chunk of impure reasoning and restore the proper language. It could be in the end a non-effective affair, as non-effective as the completeness theorem itself.

One could, however, object that in the case of a very informal proof the mathematician has not proved the theorem, but has at most shown that there exists a proof. This description can be accepted, together with the related one of the mathematician as a *meta-mathematician.* First of all such description is realistic; the mathematician acts as a general scientist, often relying on consequences and side effects to show that a theorem must hold. Secondly, it explains the fact that mathematicians prefer to work with stronger theories, rather than with logically strictly sufficient assumptions. They like set theory, because its language is less mathematical and more akin to natural ones (some say it is a logical language). This preference tells us a great deal about the meta-mathematical attitudes of mathematicians. Greater ease of finding proofs, even shorter ones, in stronger theories can be formalized by reflection principles. Probabilistic proofs of the existence of a proof, such as the zero-knowledge proofs of S. Micali and S. Goldwasser, and interactive protocols could fit into this meta-mathematical frame (see the expository papers of Buhler (1986) on zero-knowledge proofs and of Cipra (1992) on the work of L. Babai's team).

But this is an extreme and partial position: if actual reasoning need not bear a resemblance to the structure of the formal proof, it *can* bear some resemblance, not only in the formal fragments always present, but in the structural development. Then we are entitled to say that the informal proof is a sketch of the formal one. The similarity of informal and formal proofs is perhaps still more surprising then their divergence. The formal proof in itself is not a discourse, let alone a reasoning; it is a chain of abstract elements connected together according to simple rules. But in a Paschian sense it can be seen as a schema of an actual reasoning. Trained mathematicians are able to reason formally in this way, forgetting the meaning. Now this possibility is also a consequence of the completeness theorem, not of what it says, but of its being provable at all, or of the way it is proved. It is a consequence of the fact that the rules for which completeness holds and by means of which we enchain formal derivations have been isolated, or abstracted from a long history of argumentative practices.

The rootedness of completeness in human history and nature might enhance our confidence in it, but does not make its truth less amazing. We can probably accept that it is possible to build (or to prove that there exists) a finite structure which encodes the information relative to the fact that a logical truth holds; but that this code, compressing information concerning the possible interpretations and what happens in them to our formal sentences, should have a parallel in what goes on in our brains (or is reported as such by the language) is a new version of pre-established harmony that does not cease to surprise.

Could it be that the rules are laws of thought? It is unlikely, and not even meaningful; we do not know what a law of thought could possibly be. In any case we do not have a completeness theorem for this notion, but only for the extensional one.

4 The psychology of mathematical thinking

The question however of how mathematicians can reason formally deserves investigation. The psychological literature does not help us much; scholars are still debating pseudo-foundational issues, such as whether there are mental laws or not; for two competing approaches, see Johnson-Laird (1983) and Rips (1994). To reason formally means combining formulae and at the same time capturing the notion of logical truth. How do we perceive logical truth (in itself, apart from and preliminary to the proof), or how do we arrive at it?

When a mathematician begins his argument (let us not consider for now false starts) to reach B, or some still unknown or confused thesis, he thinks of A, or better, since one does not think of a formula, he builds an image, a mental model of A. In the usual jargon, he considers a structure in which A is true. The syntactic elements of A suggest the relevant features to visualize (relations, functions), and their mutual connections. After setting the stage, the mathematician then goes on describing what he sees as being true in the structure, possibly after active intervention on his part to disclose some hidden properties.

If the conclusion B is acceptable as a logical conclusion, however, there must be something else; if a statement is true in a structure it is not a theorem of the theory having that structure as a model, but for special cases of complete theories—a notion mathematicians are seldom aware of, and in any case complete theories are rare and exceptional.

In the end, conclusion B is true not only in the envisaged structure, but in all those satisfying A. People say that they perceive truth in a structure, while what they are doing is more (or less if you prefer): they are perceiving logical truth. It must be the case that the mathematician is using by experience only those properties of the structure which are actually shared by all models. This is more easily said than done, or explained how it is done. But on second thoughts it could not be that mysterious. This notable performance could depend on the nature of seeing.

Seeing is not a passive action, not even at the physiological level, where neurophysiology shows that it is heavily theory driven, the more so for abstract objects. To see a mathematical object we have to know at least that we want to do mathematics. We do not see a triangle, even if we have three points in front of us, unless we are disposed to do geometry. Otherwise, we could also see a graph. What we see depends on the problem and on the theory we are developing.

In mathematics, seeing does not consist in seeing a drawing or a particular concrete set. Mathematicians see *structure*, not *a* structure. Truth in a structure, which figures so prominently in the mathematical discourse, would not be a problematic notion, but for the fact that structures are not there to be seen. When people say that they are describing a structure they are not describing what they see, even with the mind's eyes, as a pre-existing reality. They are expounding a (form of) description and at the same time seeing what the description purports to be about.

This means that mathematicians see only things or properties describable in words. Sometimes words may be lacking and it takes some time to pass from the sensation of grasping something important to a descriptive statements, with new words which have to be invented and mastered—mathematics is the field where more new words keep being invented.

Mathematicians know, from the axiomatic method (indeed, from the completeness theorem itself), that when they describe a concept in words, they always end up with an incomplete description. So the image evoked by A is never unique; as far as describable properties are concerned, it is the common image of a class of structures. And truth in the imagined structure means automatically truth in a lot of structures falling under the same description.

Different models may differ for negligible particulars, or else for momentous ones, but such as not being expressible in words (of the language used). A good example is that of non-standard models of arithmetic; we have an overall image of them, but we cannot say from within how they differ from the standard one, at least using statements of first-order arithmetic. We cannot really imagine a model being non-standard if we are working in it according to arithmetical axioms; we have to focus our attention on an explicit non-standard element. When we are really able to say, not just to see, in the language of arithmetic that we have two different models, then we enter the realm of independence proofs.

So a proof begins by stating the hypotheses and thus seeing some structure. Sometimes the proofs end in the same way, since the conclusion is easily seen in the picture, or with a few added elements. It is the case of Euclid's proofs consisting of just one word: Look! But also in this case there is some logic involved. The very possibility of forming a mental model of the assumptions depends on the assumptions being consistent.

The last statement appears to be contradicted by proofs by contradiction, or in some cases of false starts: here one states a (later to be shown) impossible hypothesis, but seems to be able to visualize it, as in a normal proof, even to draw (part of) a picture (later to be shown incompletable). The reason is that, in stating a hypothesis, one is not really seeing, but just carefully considering a description and preparing to look at the consequences of the description. The image is so to speak held over; the picture is blurred in critical points. If the description turns out to be consistent, then one can safely see the related image. The explanation is given again by the completeness theorem, in the equivalent form of the model existence theorem: if a set of sentences is syntactically consistent, then it has a model. If the description turns out to be inconsistent, the image was an illusion, not unlike those we have in visual perception. But contradictions are impossible to visualize, they can only be stated in words; this is one of the difficulties of psychological theories of mental models which are presented as alternative to logic.

Seeing and deducing are tied together. If the conclusion is not immediately evident in the model associated to the assumptions, then the proof develops the description in a discursive way, enlarging the view, adding details—always suggested by the syntactic structure of drawn consequences—until the conclusion is seen by everybody.

A good proof is going to give *the reason why* the conclusion depends on the assumptions. Different proofs give different reasons. The reader is invited to reflect on different proofs of his favourite theorem; mine in the present context is Euler's theorem on connected simple graphs: such a graph has an Euler circuit if and only if the degree of every vertex is even. The degree of a vertex is the number of different edges incident on it. A circuit is a path whose last vertex coincides with the first one. An Euler circuit is a circuit that contains each edge of the graph just once; it is only an Euler path if it does not contain all edges. Call a connected simple graph an Euler graph if every vertex has an even degree. The necessity of the condition is obvious; to some also the sufficiency; this is probably why Euler did not prove the theorem; see Fowler (1988) for further information.

One proof, reported in Galovich (1989, pp. 303–13), is roughly as follows. Suppose that there exists an Euler graph without Euler circuits, and let G be one such graph with a minimum number of vertices. Consider in G a maximal circuit with no repeated edges σ, which cannot be an Euler circuit. Then $G-\sigma$ is obviously defined and easily seen to be still an Euler graph, with a lesser number of vertices; it has a connected subgraph with a vertex in common with σ; here there is an Euler circuit; by linking it with σ, one contradicts the maximality of σ. The details can all be filled in and are not important; the point is that this proof does not show how the hypothesis of the even number of edges works, but it must be verified in order to apply the induction hypothesis.

There are algorithms to define Euler circuits in Euler graphs; one such algorithm is that of Fleury, which says that each move in the stepwise construction of the circuit is indifferent, so long as you do not blatantly blunder: if vertex A has been reached, take as next edge any edge incident on A, unless by doing so you put yourself in the impossibility of continuing, since the graph remaining after taking away the path so far constructed turns out to be disconnected. The above sketched proof could not be transformed into a proof of the correctness of the algorithm; for this, one needs a proof based on the fact that with an even number of edges one can always enter any vertex from an edge and leave by another, or conversely.

The proof might run as follows, again by induction. Given an Euler graph G with $n+1$ vertices, let us consider a vertex A; A could be the vertex reached at any stage in the construction of the circuit. If there are only two edges leaving from A to the same B, then it is easily seen how they can be eliminated, reducing G to a graph with n vertices. If there are two edges b and c connecting A to B and C, then there are two cases. Suppose that B and C are connected, and π is an Euler path connecting them; then by subtracting from G the cycle formed by π and b and c one can apply the induction hypothesis; this means, however, that, if we enter A through b, then by going to C through c and coming back to B through π we can resume the thread in the remaining part of the graph.

If B and C are not connected, then we get a contradiction; by eliminating A, b, and c and identifying B and C we obtain an Euler graph; but leaving B through one of the old edges incident on B—an odd number of them—we can come back to it only through one of the same, and similarly if leaving it through an old edge incident on C; count of parity shows that this is impossible. The case is not difficult but it is instructive; it is an example of what we said above on the visualization of impossible situations; the drawing one contemplates is the same as before, with parts of the graph shown under B and C, some edges and some dots and dashes; but it does not exist. There are other cases to consider—more than two edges incident on A—but they are not meaningful.

Both proofs are almost formal, in the sense that they are written in the language of the problem; they can be formalized by adding details without changing their overall structure; but they also show, in different ways—in our opinion one more than the other—why the conclusion must hold in the envisaged Euler graph.

5 Conclusion

We have been inviting you to reflect that the old-fashioned features of the logical notion of consequence are in a sense preserved in the formal one, or at least they are not incompatible with it. We want to check validity in all interpretation; we know that a formal proof encodes by syntactic structures all the relevant information concerning possible interpretations; a proof is a link between assumptions and thesis realized by syntactic transformations; to show that it is possible to transform the assumptions in the thesis and to point to the direction, if not actually doing all the steps, of the transformation we give compelling and convincing reasons that a particular use of the premises will get the conclusion (that the latter is contained in them, as it used to be said in the logical tradition). Very often, these reasons consist in just seeing how the premises get transformed.

At the end, we must anticipate an obvious rejoinder: all this talk of theorems in the light of the completeness theorem should apply self-referentially to the completeness theorem itself. But it does, with some peculiarities.

As a mathematical property, completeness is best formulated and proved in set theory, having to deal with languages and interpretations. It has a proof—indeed several proofs—not too long, quite manageable; this does not mean that is absolutely certain; proofs serve to show the reasons why we should accept a theorem. As a set-theoretical notion, it shares in principle the possible interpretations of the set concept, which admittedly are not easy to define. Among them there is the anthropological one referring to the languages we use and the thoughts we think.

On the other hand, being equivalent to the Boolean prime ideal theorem, and only a little less strong than the axiom of choice, the nature of completeness is more that of an axiom than of a theorem. We have not been discussing axioms, but this much can be said, if not exhaustively: that axioms are accepted for their usefulness, and that they usually say that we can reason in a certain way (inductively, or by algebraic manipulations, or by appealing to geometrical concepts such as continuity). The completeness theorem partakes of all the mathematical arguments in favour of the Boolean prime ideal theorem; moreover, it has a great deal to say about reasoning.

Bibliography

Appel, K. and Haken, W. (1986). The four color proof suffices. *The Mathematical Intelligencer*, **8**, 10–20.

Buhler, J. (1986). Zero-knowledge proofs. *Focus*, Newsletter of the Mathematical Association of America, **6**, no. 5, October, p. 1.

Cipra, B. A. (1992). Theoretical computer scientists develop transparent proof techniques. *SIAM News*, **25**, May, p. 1.

Davis, P. and Hersh, M. (1980). *The mathematical experience.* Birkhäuser, Boston.

Detlefsen, M. (ed.) (1992). *Proof and knowledge in mathematics.* Routledge, London.

Fowler, P. A. (1988). The Königsberg bridges—250 years later. *The American Mathematical Monthly*, **95**, 42–3.

Galovich, S. (1989). *Introduction to mathematical structures.* Harcourt Brace Jovanovich, San Diego.

Hersh, R. (1979). Some proposals for reviving the philosophy of mathematics. *Advances in Mathematics*, **31**, 31–50.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge University Press.

Kreisel, G. (1967). Informal rigour and completeness proofs. In *Problems in the philosophy of mathematics* (ed. I. Lakatos), pp. 138–71. North-Holland, Amsterdam.

Lolli, G. (1987). *La macchina e le dimostrazioni.* Il Mulino, Bologna.

Lolli, G. (1995). *Completeness.* Associazone Italiana Logica e Applicazione, Milan. Preprint.

MacKenzie, D. (1995). The automation of proof: a historical and sociological exploration. *IEEE Annals of the History of Computing*, **17**, 7–29.

Maddy, P. (1998*c*). How to be a naturalist about mathematics. *This volume*, 161–80.

Manin, Yu. I. (1998). Truth, rigour, and commonsense. *This volume*, 147–59.

Pasch, M. (1882). *Vorlesungen über neuere Geometrie.* Teubner, Leipzig.

Penrose, R. (1994). *Shadows of the mind.* Oxford University Press.

Putnam, H. (1995). Review of Penrose (1994). *Bull. American Math. Soc.* (N.S.), **32**, 370–3.

Quine, W. V. (1970). *Philosophy of logic.* Prentice-Hall, Englewood Cliffs, New Jersey.

Rips, L. J. (1994). *The psychology of proof.* The MIT Press, Cambridge, Massachusetts.

Swart, E. R. (1980). The philosophical implications of the four-color problem. *American Math. Monthly*, **87**, 697–707.

Dipartimento di Informatica
Università di Torino
Italy
email: gabriele@di.unito.it

7

Mathematics as language

Edward G. Effros

1 Introduction

According to an anecdote popular among physicists, it is easy to tell when you are speaking to a mathematician: if you should ask him 'do you or do you not own an umbrella', it is likely that after some thought he will answer 'yes'.

Ever since mathematics was codified in terms of symbolic logic in the early part of this century, it has been understood that *all* of mathematical discourse consists of tautologies of the above variety. Owing to the seemingly absolute precision of axiomatic deduction, few mathematicians are concerned with the *grammatical* truth of their subject. Thus despite the fact that they recognize that the axioms of logic are necessarily incomplete, and even possibly inconsistent, most mathematicians are confident that any future difficulties will yield to technical adjustments. This complacency has been reinforced by the remarkable progress that logicians have made in delineating the limits of the axiomatic method.

By contrast, in the broader context of *meaning*, questions of truth are of growing concern to a wide range of mathematicians. In a trend that is particularly evident in the United States, a number of individuals both within and outside the profession have questioned the methods and even the purposes of mathematics. Many of my colleagues were dismayed with the recent publication in *Scientific American* of an article entitled "The death of proof" predicting a major shift in the way mathematics would be done (Horgan 1993). This view was also promoted by D. Zeilberger (1993). In that article he predicted that with the ever-increasing use of computers, the very nature of mathematics will change. He went on to prophesy:

> Although there will always be a small group of 'rigorous' old-style mathematicians..., they may be viewed by future mainstream mathematicians as a fringe sect of harmless eccentrics...

With the weakening of the mathematical establishment in Eastern Europe, mathematics has become more monolithic. As a result, what is fashionable in the United States is more likely to take root elsewhere, and these ideas must be taken seriously by mathematicians everywhere. I believe that these challenges to the discipline are due in part to a fundamental misunderstanding of the nature of mathematical thought. This meeting has provided a timely opportunity to explain to a wider audience what many mathematicians believe constitutes the 'truth' that may be found in mathematics, and the manner in which we believe that it is threatened by recent developments.

This chapter begins with a brief explanation of why I believe that mathematics is in essence a language. Next I summarize some of the more disturbing attempts that are being made to change the ways in which mathematics is taught and used. In subsequent sections I will try to show that these new methods are likely to damage the basic content of mathematics. I conclude with an illustration of one of the many ways in which mathematics is continuing to evolve, by considering recent attempts to quantize mathematics.

2 Mathematics is most valued as a language

The recent solution of Fermat's problem undoubtedly came as a blow to a large number of amateur mathematicians. Fortunately, there is no reason for despair: many problems which are even more important remain open. Amateurs and professionals alike cherish the idea that they might some day solve a famous open problem. The formulation and solution of problems has provided what is regarded by many as the most characteristic feature of the subject. It must be stressed, however, that the primary purpose of problem solving is to facilitate the discovery of new mathematical concepts, and to gauge the success of these methods. Those involved in the solution of Fermat's problem have been quick to point out that the new ideas used to solve the problem are much more important than the result itself.

Much of mathematical research is devoted to developing new machinery without specific problems in mind. The success of modern mathematics is in large part due to the axiomatization and elaboration of such notions as connectivity, symmetry, smoothness, and infinite-dimensional analysis. These represent fundamental developments in the *language* of mathematics, and it is precisely these conceptual advances that have proved to be so essential to modern science.

Our premise is that mathematics is a language since it provides both a conveyance for and a substantiation of our thoughts. It is that aspect of mathematics that explains the key role it plays in modern science. This is well illustrated by modern physics. One need only briefly examine current journals to appreciate the extent to which modern mathematical concepts such as curvature and connectivity have been adopted by the physicists. Although they are not as concerned with mathematical precision, *they are fully aware that one cannot effectively use the modern mathematical tools without a deep understanding of the associated deductive machinery.* The success of modern physics is in no small part a consequence of the mathematical language they have at their disposal.

3 Three challenges to mathematics

Two of the recent criticisms of traditional mathematics are pedagogical in nature. Mathematical research is inextricably linked with mathematical education. The imperative to pass on the theory to succeeding generations of young mathematicians is particularly strong in mathematics, since young investigators play a central role in the subject. Furthermore, teaching mathematics requires that researchers reformulate and clarify their work, and to a large extent it is this process that has consolidated our mathematical knowledge. It is thus not surprising that challenges to the subject are often first reflected in attempts to change mathematical instruction.

It is argued that mathematical fluency is being over-emphasized in the schools.

Quoting from two documents widely circulated by the National Research Council:

> Since few arithmetic calculations are done most efficiently using paper and pencil, the level of arithmetic skill that is the current goal in most elementary school class rooms is far in excess of what is needed for tomorrow's society. (Research Council 1989*b*)

and

> Weakness in algebraic skills need no longer prevent students from understanding ideas in more advanced mathematics. Just as computerized spelling checkers permit writers to express ideas without the psychological block of terrible spelling, so will the new calculators enable motivated students who are weak in algebra or trigonometry to persevere in calculus or statistics. The argument has been made that with the availability of calculators and computers, students with only a minimum background in the basic techniques of algebra or even of arithmetic should be able to take suitably designed calculus courses. (Research Council 1989*a*)

Many of the new textbooks have taken the solution of 'problems arising in the real world' to be the primary purpose and motivation of mathematics. These texts base their approach on communicating a 'feeling' for mathematical concepts based on pictures, computer experiments, and qualitative arguments.

Proofs are indeed dead in secondary and even in college education in the United States.

In a dramatic change of curriculum, the method of proof has been virtually eliminated from the high schools and the elementary college courses. Recently Steen, a well-known figure in mathematical education, examined a large sample of calculus examinations given in a number of colleges. He reported:

> You find very rarely—only 1 problem out of 20 examinations—the kind of question that used to be very common 20 to 30 years ago: 'State and prove ...' problems dealing with the theory of calculus or with rigorous calculus have simply vanished from American calculus examinations. (Steen 1994)

This development was in part a response to an earlier misguided attempt to introduce formal set-theoretic methods into elementary school mathematics. Despite the warnings of a number of mathematicians that it was a pedagogical error to introduce these notions to primary school children, the 'new mathematics' swept the nation in the sixties and seventies, with often catastrophic consequences.

The elimination of deductive methods is particularly evident in some of the latest 'reform' calculus texts, in which some of the fundamental deductive building blocks of the subject, including the mean value theorem, have been dropped from the syllabus. Turning to a striking quotation from a recent guideline for educators,

> ... secondary school is [not] the place for students to learn to write rigorous, formal mathematical proofs. That place is in upper division courses in college.

This was quoted in (Steen 1994)—the latter contains an excellent analysis of the current fashion to de-emphasize proofs.

It is being claimed that computers have rendered many of the methods of mathematics obsolete.

It is indeed the case that many of the concepts of mathematics were developed precisely because we did not have efficient methods of computation in the past. Thus it is claimed that the mental gymnastics that these methods required now serve little purpose. It has been suggested that mathematicians would better spend their time if they approached mathematical problems from an empirical point of view. With the advent of computer technology, we can now perform 'experiments' in mathematics which can indicate what is 'likely' to be the case. Some even claim that our search for the certainties of logical deduction no longer serves any useful purpose.

4 Language and fluency

The current educational reform movement is but one in a long series of experiments that have been attempted in the United States. These have been prompted by an increasing feeling of despair throughout the academic community. The decline in the preparation of American students in virtually all forms of intellectual endeavor is universally acknowledged (see, e.g., (Chira 1991)). The impact of this deterioration has been particularly severe in mathematics, since the discipline requires many years of preparation. As in dance or music, most students who have not received a reasonable mathematical background in their early years are irreparably damaged by the time they reach college.

Reviewing the first two challenges considered in the previous section, we note that they are both concerned with the *language* of mathematics. The abilities to do unassisted computations and frame precise proofs are the two most important components of mathematical *fluency*. Thus the proponents of these changes are advocating that we dilute our attempts to teach our students how to use a *coherent* system of mathematics.

The argument that we should instead concentrate on teaching 'problem solving' methods represents a basic misunderstanding of the purpose of mathematical education. We do not include algebra in the high school curriculum in order to enable students to solve 'word problems'. Although these exercises are essential for motivating and reinforcing the algebraic techniques, they are not particularly important in themselves. The reason that we teach algebra is quite simply that it is an essential part of the language of modern science and engineering, and that without it, students cannot enter those fields.

Turning to a related aspect of the 'reform' movement, it should be noted that the simple-minded problems used in algebra can inevitably be solved *without* algebra by using graphical and computer methods. These methods are elegant and often require originality on the part of the students. But it is a mistake to over-emphasize these approaches. Although we must of course encourage original thinking on the part of our students, once again, the most invaluable skill that we can give to our students is mathematical fluency—the ability to speak the language. Without that facility, each student will be forced to continually 'reinvent the wheel'.

The pedagogical principles underlying mathematics instruction are quite similar to those used in language instruction. One need only consider what it would be like to teach a French literature class to students who did not have a basic knowledge of French vocabulary and grammar. There is no way that a student can understand a French text if he can only stumble over the simplest words. But it is just this point that is not acknowledged in the current mathematical texts. Students who have not fully mastered the one-place multiplication table (which is the case for many of our college students today) will learn little from the pictures and calculators that are supposed to make the subject more accessible.

Just as it would be ridiculous to claim that hand-held translators provide a pivotal technology for teaching language, the same is true for calculators in mathematics. There is no question that these devices provide worthwhile new approaches to the subject. However, it does not seem to be understood that *elementary calculations should never be done on a calculator.* Imagine the situation of a student who tries to understand algebra if he does not know arithmetic. When confronted with the problem of computing $(x^8)^7$, he will be disconcerted by the fact that he has to first calculate 8×7. A lack of fluency is even more crippling when one tries to understand calculus. The fact that one will soon be able to factor polynomials on a hand-held calculator is completely irrelevant to the requirement that our students must be able to *instantly* factor an expression such as $x^2 - 9$.

The current stress on calculators and computers is a distraction from the primary educational issues. The most immediate cause for the illiteracy and 'innumeracy' of our students is the simple fact that they are devoting only a minimal amount of time to their studies (see, e.g., (Chira 1991)). These problems are rooted in our societal decline, and they must be addressed by political action to change the attitudes of the citizens toward education.

Regarding the second challenge, the decision to delay mathematical deduction until upper division mathematics has proved to be disastrous. It is becoming increasingly clear that this is a problem that cannot be solved by remedial courses—it is just too late for most of our students. We have heard from colleagues across the country that it is now virtually impossible to teach basic courses such as Real Variable theory to our undergraduates. As a result, our students are grossly unprepared for graduate school. At the graduate level we now find that many (most?) of our graduate students are simply unable to prove theorems.

The most pernicious effect of these developments has been the increasingly popular claim that the basic methods of mathematics, and in particular, the method of deduction, are not important to mathematics or to society at large. These attitudes are related to the third challenge to mathematics, which we will consider in the next section.

5 Computers and mathematics

Chess provides a microcosm of the world of mathematics. It is now not unusual for a computer to defeat a grandmaster in a chess game. With the exponential increase in computational speed that has continued unabated for the last forty years, it seems only a matter of a decade or so before a desktop, or perhaps even a hand-held computer, will be able to overcome any human being. If and when this happens, will the world of chess also be reduced to 'a fringe sect of harmless eccentrics'?

Some argue that the demise of chess is unlikely. Pointing out the parallel in athletics, they note that runners have not been discouraged by the invention of bicycles or cars. We might therefore expect chess players to be sufficiently motivated by the intellectual challenge of the game. Furthermore, it seems probable that, with time, computers will become a powerful tool for understanding the nature of the game.

Unfortunately, there is evidence that this is not a valid comparison. It is a sad truth that most of us find mental exertion much more painful than physical effort. Many regard unnecessary mental gymnastics as more a sign of masochism than of self-discipline. There may in fact be a physiological mechanism underlying this phenomenon. We are told that the feeling of well-being that is experienced by an athlete is mediated by a chemical system of endorphins and their receptors. As a result, physical activities such as running can even become addictive. Although intellectual success can lead to increased self-esteem, and for some can become an obsession, there does not seem to be a direct biological mechanism that is involved. There is little evidence that we have a 'mathematics pleasure receptor' in the brain.

The possible effects of computers on chess were recently examined by Mark Saylor (1997), a US national champion, in a newspaper report on an upcoming match between Kasparov and IBM's program "Deep Blue". Quoting an international master (Peters), he wrote

> A computer victory would devastate chess, though the impact would not be felt for a generation. ... I fear that if computers are perceived to play chess better than humans can, then new future Kasparovs won't study chess.... Chess requires years of study to reach the grandmaster level. If computers are perceived to be the best, the people with the biggest egos won't take it up. We'll still have chess and still have top players, but not Kasparovs.

Regarding mathematics, it is worth recalling how pocket calculators changed the public's view of 'mental computers'. Fifty years ago there was a select group of individuals who devoted themselves to performing arithmetical calculations in their heads. Their remarkable feats attracted considerable admiration from the public. Today calculators and computers serve as the 'great equalizer'. Individuals who enjoy calculating in their heads are regarded as more peculiar than talented.

Computers have unquestionably rendered certain techniques of mathematics obsolete. Virtually no one uses logarithms or slide-rules to do arithmetical calculations any more. A more instructive example, however, is associated with the problem of *integration*, i.e., the calculation of continuous sums. This constitutes the computational heart of the physical sciences, since physical laws are generally formulated in terms of 'infinitesimal quantities' which must be integrated to make predictions.

Integration problems were first systematically studied by the Greek mathematicians, who tried to determine the areas of simple geometric figures. Of course it had been understood long before that time that one could approximate an area by trying to pave the interior with a collection of small rectangles. But it was the Greek mathematicians who made the first attempt to develop an *exact* theory of areas. They recognized that the areas of 'ideal' figures in classical geometry, such as the interiors of circles, are universally fixed, and that one might try to *relate* these constants. Although the Greek mathematicians were able to find a few such relations, it was not until the seventeenth century that a general approach was discovered. The fundamental theorem of calculus provided a magical key for integration, and a significant portion of modern mathematics grew out of the need to perfect this method.

It has been proposed that if computers had been invented in the sixteenth-century, mathematicians might not have discovered the fundamental theorem of calculus. Computers make the evaluation of areas a completely routine exercise. In order to explore the implications of this idea, it is helpful to dramatize them with an imaginary scenario. The example is well-known to calculus students.

In 2195, a young applied (will there be any other type?) mathematician by the name of Dick is asked by his boss, Jane, to find the area A between the curves

$$y = -\frac{1}{1+x^2} \quad \text{and} \quad y = \frac{1}{1+x^2}$$

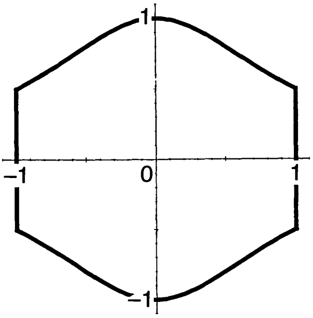


FIG. 1.

and the lines $x = -1$ and $x = 1$ (see Fig.1).

Like so many young people today, he turns to his computer and in an instant, determines that the answer is $A = 3.14159\ldots$ Being a bright fellow, it dawns on him that this is suspiciously close to π. Checking further, he discovers that his hunch is apparently true to at least a billion decimal places. Reporting the answer to his boss, he asks her why the answer should have anything to do with the area of a circle of radius 1. After informing Dick that he is wasting company time, Jane remarks that he might (on his own time) try looking in some of the old mathematics books. She is under the impression that 'people used to worry about such things'.

It is not difficult to guess the end of this story. Dick manages to locate some antique calculus books. He soon discovers that indeed, the old-timers seem to have had arcane ways of calculating areas *without a computer*, and that these techniques often enabled them to find *exact* relationships between areas of very different shapes. But calculus is difficult to learn in the twentieth century, and it will be even less accessible when it is no longer 'needed'. After some frustration with all of the abstract ideas associated with functions, anti-derivatives, and definite integrals, Dick throws up his hands with the exclamation 'How could people have wasted so much time on such irrelevant questions?'. In fact, if he were informed of the herculean efforts by mathematicians in the twentieth century to extend the fundamental theorem (see, e.g., the Atiyah–Singer theorem), Dick might be tempted to compare these efforts with the construction of the pyramids—spectacular, but useless.

If and when chess, or the fundamental theorem of calculus is abandoned, we must ask ourselves exactly what it is that we will have lost. The beauty of chess as a game is that to become adept, one must understand profound notions of *strategy and tactics* that show an uncanny resemblance to the principles of human conflict. The satisfaction of winning a game was that it seemed to demonstrate a deeper *understanding* of these concepts. But human understanding is irrelevant if *determining who wins* is our only objective. If that becomes our perception, the irrelevance of playing chess in the age of computers will be apparent. Similarly, if our only concern in mathematics is *finding the answer*, it is my feeling that the very purpose of mathematics will be lost. I find it puzzling and ironic that we are now teaching most of our students how to adjust to a world in which there is no significant mathematics.

As in the other scientific disciplines, the most valuable product of our enterprise has been discovering concepts. These concepts in turn represent extensions in our ability to use language, the mediator of human understanding.

6 The continuing evolution of mathematics

Perhaps the most intimidating aspect of mathematics is its symbols. For most young students, it is the transition from specific numbers to 'unknowns' that constitutes the most daunting feature of high school mathematics. The next trauma occurs when letters are introduced for 'complex numbers'. The biggest, and an often terminal intellectual leap, is required in calculus, in which the variables are taken to stand for unknown functions rather than individual numbers.

These successive layers of instruction reflect the evolution of mathematics. Algebra was the major tool that was missing in ancient Greece. Complex numbers first arose in an essential manner in the sixteenth century solution of algebraic equations of the third-order. We recall that the equation

$$ax^2 + bx + c = 0$$

has the solutions

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

One need not introduce complex numbers into this formula, since one can simply declare there to be no solution if $b^2 - 4ac < 0$. Turning to third-order equations, one may use graphical methods to show that the equation

$$x^3 + px + q = 0$$

has three distinct real roots for suitable values of p and q. Tartaglia and Cardano succeeded in finding a general formula for these solutions. What distinguished their result from the quadratic case was that even when one sought real roots, the formula often required that one manipulate complex quantities (see (Penrose 1994)).

With the invention of the calculus, the focus of mathematics shifted from individual numbers to functions. In this context, one is most interested in solving differential equations such as

$$y'' + ay' + b = 0\,,$$

where the letters a and b stand for given functions, and y is an *unknown function*, rather than an unknown number.

It was not until the second half of the nineteenth century that mathematicians fully realized that they were free to create all kinds of algebraic systems in which the 'variables' stood for completely new objects. It is this trend more than any other that distinguishes the essential features of modern mathematics. The key to this step forward was the axiomatic method. In each case one begins by stating the axioms (such as commutativity) that these variables and their operations must satisfy. Roughly speaking, this is analogous to the importance of generalizing Euclid's axioms in order to formulate non-Euclidean geometries.

In this century mathematics has grown in innumerable directions. I will confine this discussion to what is perhaps the most revolutionary change. In 1925 Heisenberg (1925) made the remarkable discovery that one could begin to understand particle physics provided one introduced *quantum variables.* Despite the universal acceptance of this notion in physics, most mathematicians were hesitant to explore its implications in mathematics. It is only now that these new variables have begun to have a significant impact on the subject.

In classical physics, the variables correspond to quantities that we measure, i.e., the *observables,* and their values depend on the *state* of a given experimental system. To be more accurate, an observable is the measuring device, or 'meter' for keeping track of a quantity, and the state describes how we set up the experiment. In the classical mathematical model we assume that we are given a set of states Ω and a mapping $X : \Omega \to \mathbb{R}$. The interpretation is that if we prepare the system to be in state $\omega \in \Omega$, the meter X will register the value $X(\omega)$. In order to accommodate uncertainties in the way that we set up our experiment, it is often useful to extend this notion. Instead of assuming that we are 'trying out' a particular ω, we assume that our experiment corresponds to a probability measure P on Ω. In this context X is called a *random variable.* The idea is that an experiment leads to a fluctuating reading $X(\omega)$. In this more realistic construction, the best that we can do is to describe the *variation* of $X(\omega)$, rather than assign a particular value α to X. The behavior of X is represented by a probability measure μ on $\mathbb{R}$, where the probability that a reading will occur in a given interval $B = (\alpha - \varepsilon, \alpha + \varepsilon)$ is a number $\mu(B) = P(X^{-1}(B))$ lying between 0 and 1. Turning to the algebra associated with this system, we have that the observables comprise an algebra $\mathcal{A}$ of functions X on the set Ω. It is one of the first theorems of functional analysis that, under the appropriate conditions, one can use integration to identify a probability measure P with a linear functional P on $\mathcal{A}$ satisfying $P(I) = 1$ and $P(\bar{f}f) \geq 0$ for $f \in \mathcal{A}$ (i.e., P is 'positive'). The measure μ is determined by the integral relation

$$\int t^n \,\mathrm{d}\mu(t) = P(X^n) = \int X(\omega)^n \,\mathrm{d}P(\omega),$$

or to put it another way, μ is the 'restriction' of the functional P to the subalgebra of $\mathcal{A}$ generated by X. It should be noted that we may regard μ as a probability measure on the set $X(\Omega)$. It is useful to think of the numbers in $X(\Omega)$ as the values that the meter is allowed to register. This set is also known as the *range* or the *spectrum* of the function X.

It is a more or less implicit assumption of classical physics that, in principle, experiments can be *refined*. By this we mean that, if we set up our experiment more carefully, we can decrease the variation of the readings of our observables, i.e., we can 'purify' the state. If this were not possible, the set Ω would become 'physically unanalysable'—there would be no way to physically distinguish one part of the set from another. In more technical terms, there would be no physically meaningful procedure for coordinatizing the set.

Radioactivity provided the first examples of observables that violate the classical pattern. Any attempt to segregate a sample of radium into more stable and less stable subpopulations failed. The overwhelming evidence is that radium atoms are in this sense *indistinguishable*. One cannot, for example, decrease the variation of the observable X that measures the lifetime of an atom of radium.

In order to get some feeling for quantum variables, we turn to one of the simplest of quantum experiments: the measurement of the polarization of photons. Classically, the polarization of a monochromatic beam of light parallel to the Z-axis is described by a vector $\psi = (\psi_1, \psi_2)$ in the two-dimensional complex Hilbert space $\mathbb{C}^2$. This is an elementary consequence of classical wave theory. The electric field at the plane $z = 0$ is given by

$$\begin{aligned} E_x(t) &= A\cos(\omega t + \varphi_1) = \operatorname{Re} A\mathrm{e}^{\mathrm{i}(\omega t+\varphi_1)} \\ E_y(t) &= B\cos(\omega t + \varphi_2) = \operatorname{Re} B\mathrm{e}^{\mathrm{i}(\omega t+\varphi_2)} \end{aligned}$$

or, in other words, letting $\mathbf{E}(t)=(E_x(t), E_y(t))$, we have that

$$\mathbf{E}(t) = \operatorname{Re} \psi \mathrm{e}^{\mathrm{i}\omega t}$$

where $\psi = (A\mathrm{e}^{\mathrm{i}\varphi_1}, B\mathrm{e}^{\mathrm{i}\varphi_2})$. The vector ψ has length $\|\psi\| = \sqrt{A^2+B^2}$, where $I = \|\psi\|^2$ is just the *intensity* of the light beam. For simplicity, let us restrict our attention to the case that ψ is a vector in $\mathbb{R}^2$. Physically this corresponds to a beam that is polarized in a plane determined by a vector ψ in the (X, Y)-plane (complex vectors include circularly polarized waves). We may also use a *unit* vector θ in the (X, Y)-plane to indicate the direction of a polaroid filter placed perpendicular to the beam. The key calculation is that when a beam with state ψ passes through a polaroid filter whose state of polarization is θ, the resulting beam has the state vector $(\psi \cdot \theta)\ \theta$, and in particular it has intensity $|(\psi \cdot \theta)|^2$.

In the first decade of this century, it was discovered that a light beam was composed of particles called photons. It was through the attempt to understand how these photons 'conspire' to give the beam its physical properties that quantum mechanics first arose. In this context, Heisenberg proposed that the state ψ of the beam is the *only* physically meaningful quantity that can be attributed to each of its photons. Furthermore, since the intensity $I = \|\psi\|^2$ simply corresponds to the *number* of photons in the beam, the state of an individual photon is completely described by the normalized unit vector, i.e., we may assume that $\|\psi\| = 1$. An individual photon does not have any further properties that can be used to predict its polarization behavior. Thus when a photon of polarization ψ

passes through a θ polarized filter, all that can be said is that with probability $|(\psi\cdot\theta)|^2$ it will pass through and come out with the state vector θ, and otherwise it will be absorbed.

In a quantum system, observables depend upon the states in an entirely new manner. In particular they cannot be modelled by functions or by classical random variables, but instead one must use *matrices* for their description. From the point of view of mathematics, this is not a completely mysterious development. In what must be regarded as one of the most spectacular convergences of physical theory and the axiomatic technique in mathematics, Gelfand and Naimark showed that complex numbers, functions, and matrices are all examples of a single simple notion: an element of a *C^*-algebra* (Gelfand and Naimark 1943). This theory has provided us with *the most natural axiomatic extension of the number concept.* As in the case of numbers, the elements of a C^*-algebra $\mathcal{A}$ can be added together and multiplied, and there is a $*$-operation $a \mapsto a^*$ which is analogous to the conjugation operation $\alpha \mapsto \overline{\alpha}$ for complex numbers. (See also (Jones 1998).) We may also assume that $\mathcal{A}$ has a multiplicative identity I. The analogue of the real numbers are the *self-adjoint elements*, i.e., the elements a in $\mathcal{A}$ for which $a^* = a$. Furthermore there is a complete norm $a \mapsto \|a\|$ which, for the complex numbers, is just the absolute value operation $\alpha \mapsto |\alpha|$. Over and above the usual Banach $*$-algebraic axioms we need assume only one more condition, namely that $\|a^*a\| = \|a\|^2$.

Restricting to the bounded case, a *quantum variable* is just a self-adjoint element a of a C^*-algebra $\mathcal{A}$. The *quantum states* are just the linear functionals p on $\mathcal{A}$ satisfying $p(I) = 1$ and $p(a^*a) \geq 0$ for all a in $\mathcal{A}$. To see how quantum states typically arise, we note that, if $\mathcal{A}$ is the algebra $M_2(\mathbb{C})$ of 2×2 matrices, and we write a for a matrix

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix},$$

then a unit vector $\psi = (\psi_1, \psi_2)$ in $\mathbb{C}^2$ determines a state p_ψ on $\mathcal{A}$ by the relation

$$p_\psi(a) = a\psi\cdot\psi = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}\begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}\cdot\begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = \sum_{i,j=1}^{2} \alpha_{ij}\psi_i\overline{\psi_j}\,.$$

Returning to polarized photons, a simple example of an observable is a device that will check to see whether a photon is polarized in the direction of a unit vector θ in the (X,Y)-plane. To be more precise, we pass the photon through a birefringent crystal which splits the beam into a beam of θ polarized light and a beam of light polarized in the perpendicular direction. We then use photomultiplier tubes to check which path has been taken by a given photon. Finally we attach these to a meter $a = a_\theta$ which registers 1 if the photon is θ polarized and -1 if it is polarized in the perpendicular direction. As in the case of classical random variables, we cannot associate a specific number with this procedure since the outcome will vary unpredictably. Rather it is determined by a probability measure on the allowed values or *spectrum* of the variable a (in this case -1 and 1).

Assuming that our birefringent crystal has the appropriate axis along the X-axis, i.e., if we let $\theta = (1,0)$, the observable $a = a_\theta$ may be identified with the self-adjoint 2×2 matrix

$$a = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and a state with unit vector ψ determines a probability measure μ on $\{-1, 1\}$ by the relation

$$\int t^n \mathrm{d}\mu(t) = a^n\psi \cdot \psi = p_\psi(a^n).$$

Putting it another way, μ is the restriction of the state p_ψ to the subalgebra of $M_2(\mathbb{C})$ generated by the observable a. Using this 'calculus', it is easy to check that if a photon comes from a beam of light polarized at an angle of 45° to θ will produce 1's and -1's with equal probability, i.e., μ will simply be the measure which assigns $1/2$ to the values 1 and -1.

Without going into any further details, we note that purification is excluded in this simple model. The 'physically pure' state of polarization ψ necessarily determines a probabilistic state on the allowed values $\{-1, 1\}$. This is not a defect of the model—the implied variation of the observable a corresponds to the experimental fact that ψ photons are indistinguishable! The limits of our physical knowledge are clearly reflected by the mathematical structure. On the other hand, this example provides a typical illustration of how precise mathematical models help one to abandon notions that are not physically meaningful. In this case, it is predictability that must be discarded.

There is a common misunderstanding among mathematicians, and even some physicists, that quantum mechanics is an incomplete theory that provides probabilistic distributions because we have not developed sufficiently sensitive instruments. Certainly there will be remarkable discoveries in the future, but most physicists are convinced that along with such classical notions as energy and momentum, or the relativistic understanding of time and space, quantum variables will always play a central role in our understanding of the universe.

A seemingly endless series of books has appeared in which various attempts have been made to explain quantum theory to non-specialists. A most compelling introduction to the subject has recently appeared in a rather speculative book by Penrose (1994). In his beautiful exposition, Penrose emphasized that the quantum-theoretic view of the world is not more *restrictive* than the classical picture, it is *simply different.* Quantum objects have remarkable non-classical properties that enable one to observe things in manners that are classically 'inconceivable'. Specifically, *photons have the 'impossible' ability to check things out without actually 'being there'.* With the advent of lasers and other devices, it is now possible to witness such physically paradoxical phenomena on a macroscopic level (see Penrose's discussion of the 'bomb tester' (Penrose 1994, Chapter V) and (Kwiat *et al.* 1996)).

As in other physical theories, the elegance and ultimately the simplicity of the mathematical framework has given additional credence to quantum theory. But the reverse is also true. It is unlikely that the theory of quantum variables in mathematics would have ever been conceived without the stimulus of the physicists. The study of quantum variables now constitutes one of the most exciting and vital portions of modern mathematics, but that is a story that must be told elsewhere (see, e.g., (Connes 1991)).

7 Conclusion

As we hope is apparent from the above discussion, mathematics is continuing to develop in exciting new directions. This is mathematics at its best, in which completely new ways of thinking are being developed. The mathematical success of this program will ultimately be judged by the depth of the rapidly increasing opus of new theorems that have been proved using these new techniques.

We must make it clear to our students that the purpose of mathematics is to expand our powers of thought, and not just to 'get an answer'. As in any language, we must first acquire an instinctual knowledge of the basic vocabulary, before we attempt to use the language effectively. We are doing our students a grave disservice by de-emphasizing fluency, and we are doing the profession a disservice by minimizing the importance of conceptual thought.

Science has reached the pinnacle of success in this century. The prospects for the twenty-first century are clouded. It remains to be seen whether there is a sufficient body of mathematicians who will defend their discipline.

Bibliography

Andrews, G. (1994). The death of proof? Semi-rigorous mathematics? You've got to be kidding. *The Mathematical Intelligencer*, **16** (4), 16–18.

Chira, S. (1991). The big test. The week in review. *The New York Times,* 24 March.

Connes, A. (1994). *Noncommutative geometry.* Academic Press, New York.

Gelfand, I. and Naimark, M. A. (1943). On the embedding of normed rings into the ring of operators. *Mat. Sbornik*, **12**, 197–213.

Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik*, **34**, 879.

Horgan, J. (1993). The death of proof. *Scientific American*, **269**, 74–103.

Jones, V. F. R. (1998). A credo of sorts. *This volume*, 203–14.

Kwiat, P., Weinfurter, H., and Zeilinger, A. (1996). Quantum seeing in the dark. *Scientific American*, **271**, 72–8.

Penrose, R. (1994). *Shadows of the mind.* Oxford University Press.

Research Council (1989*a*). *Everybody counts: a report to the nation on the future of mathematics education.* Sciences Education Board, National Research Council, National Academy Press, Washington, D. C.

Research Council (1989*b*). *Reshaping school mathematics: a philosophy and framework for curriculum, mathematics.* Sciences Education Board, National Research Council, National Academy Press, Washington, D. C.

Saylor, M. (1997). *Los Angeles Times*, 1 May.

Steen, L. A. (1987). *Calculus today, calculus for a new century.* MAA Notes, Vol. 8, Mathematical Association of America.

Wu, H. (1995). Reviews. *The Mathematical Intelligencer,* **17**, 68–75.

Zeilberger, D. (1993). Theorems for a price: tomorrow's semi-rigorous mathematical culture. *Notices American Math. Soc.*, **40**, 978–81. Reprinted in *The Mathematical Intelligencer*, **17** (4), 11–15.

Department of Mathematics
University of California
Los Angeles, CA 94720
USA
email: ege@math.ucla.edu

8

Truth, rigour, and common sense

Yu. I. Manin

To Misha Saveliev on the occasion of his 50th anniversary

The main difficulty of discussing the nature of mathematical truth in 1995 as I see it is that no new insights into it have been gained since the epoch of deep discoveries crowned by Gödel's results of the late thirties.

To avoid repetition and to enliven the discourse one can try to put the matter into a broader context and add a personal note. Both solutions tend to divert the reader's attention to vaguely related topics, and I offer my apology for choosing these dubious tactics.

This chapter is divided into three parts:

a) musings on the history of mathematics perceived as a genre of symbolic (or semiotic) games;

b) a discussion of truth and proof in the context of contemporary research (centering on a recent controversy prompted by a letter by A. Jaffe and F. Quinn (1993));

c) materials for three case studies (it being understood that the study itself will be carried out by the interested reader).

We adopt for this chapter a very naïve philosophical background.

Naïvely, a truthful statement is a statement that could be submitted to verification, and would then pass this test. Verification is a procedure involving some comparison of the statement with reality, that is, invoking an idea of *meaning.* (This applies equally well to 'evident' statements whose verification is skipped.) The reality in question can be any kind of mental construct, from freely falling bodies to transfinite cardinals. We will pass over in silence the problem of how to verify statements about transfinite cardinals, which surely will be addressed by other speakers.

The statement itself is a linguistic construct. As such, it must be grammatically correct in the first place, and meaningful in the second, before it can be submitted to a verification procedure.

Logic teaches us that certain formal constructions produce truthful statements when applied to truthful statements (syllogisms were the earliest examples). Mathematics uses such constructions recursively. All comparison with reality is relegated to comparatively scarce encounters with applications and, possibly, foundational studies. The main body of mathematical knowledge looks like a vast mental game with strict rules.

We might also contemplate the notion of truth applied, not to isolated statements, but to entities like a novel, a scientific theory, or a theological doctrine. The ideas of grammatical correctness, meaning, reality, and verification procedures acquire new dimensions, but seemingly do not lose their heuristic value. A new phenomenon is what can be called their non-locality: neither meaningfulness nor truthfulness of a theory resides entirely in its constituent statements, but rather in the whole body of the doctrine.

All the common-sense notions mentioned above have been submitted to fine theoretical analysis in many philosophical works. All of them, including the idea of reality, were also thoroughly criticized, to the extent that they were completely annihilated. One pertinent example is that of the idea of verification of a theory: it has been argued that a theory can never be verified, but only falsified.

In what follows I will try to be commonsensical and to avoid extremist views. Some truth creeps into even the wildest deconstructions of this notion, but the weaknesses of such attacks usually become apparent as soon as we start judging them by their own standards.

1 Mathematical truth in history

The modern notion of mathematical truth goes back to ancient Greece; as Bourbaki succintly puts it, 'Dépuis les Grecs, qui dit Mathématiques, dit démonstration.'

It is the demonstration that counts, which is understood as a chain of well–organized, consecutive, standard steps, not as a physical act of showing, contrary to what the etymology of the word 'demonstration' suggests.

Among other things, this means that modern mathematics is an essentially linguistic activity relying upon language, notation, and symbolic manipulation as a means of convincing even when dealing with geometric, physical *et al.* realities. Consistency of argumentation, free of contradictions and avoiding hideous gaps, plays a major role in establishing that a given utterance proves what it purports to prove. The status of the postulates P upon which the demonstration/proof of the statement S is built strictly speaking need not be discussed in mathematics, which is responsible mainly for the structure of the deduction.

This idealized image had a long pre-history, and we will try to briefly review some archaic modes of proto-mathematical behaviour.

The economic and military life of early human collectives was correlated with accounting and keeping track of food resources, the size of the tribe, the seasons, etc. Elementary arithmetic as we know it only gradually emerged as a subdialect of language supporting such activities.

Whereas the main (and for millennia, the only) form of existence of natural languages was oral speech, the oral and then written language of elementary arithmetics must have slowly crystallized from many archaic forms including counting by fingers and other body parts, collecting stones and sticks, and tying knots. (This process is now being reversed as we observe how electronic arithmetics take over the written one.)

If a mathematician is inclined to stress the 'isomorphism' of all these realizations describing the universe of natural numbers and operations on them, he must understand that this is an appalling modernization.

In terms of the classical Saussurean dichotomy Langue (as system) versus Parole (as activity), we observe a slow and difficult emergence of 'language' from 'speech', the latter involving direct manipulation of things and body parts as symbols of something else. Whatever notion of truth can be read into such activity, it must be in the final account a function of the efficiency of social behaviour supported by it. Exchange and trade furnish obvious examples. Correct counting means just exchange and profitable trade, pure and simple.

This is not however the whole story. It is important to realize that not only is it materially profitable, but virtually any form of organized behaviour can have a special meaning for a human being or a human collective. This puts archaic arithmetic on a par with rites, music and dance, and all sorts of magic. The traces of this undifferentiated perception of mathematics as a form of magic are registered quite late in the history. A person who predicts efficiently an eclipse, or an outcome of an uncertain situation, is not necessarily a sage, but more appropriately a trickster who *makes* things happen by manipulating their symbolic representations.

Many philosophers tried to demythologize the image of mathematics as predominantly intellectual activity. A. Schopenhauer for one, already in the days of modern institutionalized mathematics, wrote: 'Rechnungen haben bloß Werth für die Praxis, nicht für die Theorie. Sogar kann man sagen: wo das Rechnen anfängt, hört das Verstehen auf'. Citing this, S. Hildebrandt (1995, p.13) continues: 'Die Anbetroffenen lesen es staunend und denken sich, daß Schopenhauer schwerlich einen Blick in die Arbeiten von Euler, Lagrange oder Gauß getan haben kann'.

However, taken literally, Schopenhauer is right. Not only does computation temporarily interrupt thinking, but an ultimate justification of the act of computation is that it replaces the act of thinking (or a stage of it) by a virtually mechanical interlude, in order to support a much higher level of competence for the next act. If thought is an interiorized and tentative action, then computation is an exteriorized thought, and the degree of possible exteriorization achieved by modern computers is stunning.

In the same vein, during the previous era of biological evolution, the emergence of conscious thinking served to stop instinctive action and to replace it by planned behaviour. An animal brain calculates in order to keep the animal body alive and kicking, running, flying, seeing, hearing. A human brain does the same, and this activity is the main content of the (non-Freudian) individual subconscious which must not allow any intervention of consciousness in order not to break the complex architecture of the relevant computations. Otherwise correct (biologically optimal) results cannot be secured.

The arrival of language and consciousness in a sense allowed the human brain to elevate this unconscious computation to the level of common-sense thinking and later to the level of theoretical thinking. A price paid was a loss of spontaneity of action and emergence of less and less biological patterns of individual and collective behaviour. In short, civilization was made possible.

This complementarity of action/thought/computation tends to reproduce itself at various levels.

The new alienation of thought in computerized systems of information processing is a grotesque materialization of the (non–Jungian) collective unconscious. Its running out of control is a recurring nightmare of our society, as well as the condition of its efficient functioning.

The abstract nature of modern mathematics, understood not as its epistemological feature but as a psychological fact, supports our metaphor. The gaping abyss between the habits of our everyday thinking and the norms of mathematical reflection must remain intact if we want mathematics to fulfil its functions.

The heated battles about the foundations of mathematics which continued for several decades of this century did not resolve any of the epistemological problems under discussion. Let me remind you that at the centre of attention and criticism was Cantor's theory of infinity.

Cantor's tremendous contribution to XX$^{\text{th}}$ century mathematics was twofold. First and foremost, he introduced an extremely economical and universal language of sets which subsequently proved capable of accommodating the semantics of all actual and potential mathematical constructions. This was understood only gradually, and a full realization came only somewhere in the middle of this century. What I mean is a kind of Bourbaki picture: every single mathematical, or even metamathematical, notion, be it probability, Frobenius morphism, or a deduction rule, is an instance of a *structure* which is a construct recursively produced from initial sets with the help of a handful of primitive operations. The formal language of mathematics itself is such a structure. (Sometimes, as in categorical constructions, classes instead of sets are allowed, but from the viewpoint I am advocating here this is a minor distinction.)

I believe that Hilbert, when he spoke with prescience about 'Cantor's Paradise', had this grandiose picture in mind.

But second, Cantor produced some deep and unconventional mathematical reasonings about orders of infinity, thus spurring a long and heated controversy. As we now see it, he discovered probably the simplest imaginable and natural undecidable problem, the Continuum Hypothesis (CH). (For a penetrating discussion of the meaning of undecidability in this context, see (Gödel 1995, p. 162).)

The austere and barren world of unstructured infinite sets of various orders of magnitude undoubtedly has a magic charm of its own, and reflections about this world have in turn attracted and repelled philosophically minded mathematicians and mathematically minded philosophers for several decades. Cohen's famous proof of the consistency of the negation of CH, completing Gödel's earlier proof of the consistency of CH itself, came when the fascination with mysteries of infinity was already waning, precisely because by that time the language of sets had become the language of virtually every mathematical discourse.

Rethinking these old arguments, recalling the birth of intuitionism and constructivism, I am struck by the utterly classical mind-set of some of Cantor's critics. A considerable part of the discussion concentrated on the principles of thinking about infinite sets. The Axiom of Choice was considered basically as a wild extension of the mundane experience of picking randomly individual objects from heaps of them. Both the constructivist and intuitionist view of this picture revealed a deep emotional revulsion towards such an action involving infinite choice (in a later Essenin–Volpin decadent, ultra-intuitionistic world, even imagining finite and rather small collections of things became an unbearable strain.)

Of course, the idea of a collection of distinguishable and immutable objects belongs to layman's physics. Many actors of the great Foundations Drama seemingly were convinced that the axiomatics of Set Theory must be understood as a direct extension of this naïve physics.

The fact that even small sets of quantum objects behave quite differently was never taken in consideration. (It probably should not be.) The fact that working infinities of working mathematicians (real numbers, complex numbers, spectra of operators, etc.) were efficiently used for understanding of the real world was deemed irrelevant for foundations. (It probably is.)

In any case, the uneasiness about Cantor's arguments led Hilbert to start a deep formal study of the syntax of mathematical language (as opposed to the semantics of this language), thus preparing the ground for Tarski, Church, and Gödel (and prompting philosophical platitudes like Carnap's view of mathematics as 'systems of auxiliary statements without objects and without content', cf. (Gödel 1995, p. 335)).

What these studies taught us was a highly technical picture of the relationships between the structure of formal deductions, their naïve (or formal) set-theoretical models, and degrees of (un)solvability and (un)expressibility of the relevant precisely defined versions of mathematical truth. Popularizations ('vulgarizations') of Gödel's work rarely manage to convey the complexity of this picture, because they cannot convey the richness of its mathematical (as opposed to epistemological) context.

It is this richness that fascinates us most.

2 Truth for a working mathematician

The Bourbaki aphorism cited at the beginning of the previous section does not imply two millennia of common agreement on what constitutes a proof. Moreover, the following quotation from A. Weil's talk at the 1954 International Mathematical Congress in Amsterdam leaves an impression that the notion of 'rigorous' proof is quite recent, perhaps even due to the efforts of Bourbaki himself.

> Rigor has ceased to be thought of as a cumbersome style of formal dress that one has to wear on state occasions and discards with a

> sigh of relief as soon as one comes home. We do not ask any more whether a theorem has been rigorously proved but whether it has been proved. (Weil 1980, p. 180)

Alas, this seems to be only wishful thinking. In the individual psychological development of a mathematician and in the social history of mathematics both the understanding of what constitutes a proof and the perception of its role greatly vary.

Below I have collected a sample (A–F) of quite recent opinions of actively working mathematicians, taken from (Jaffe and Quinn 1993), and (Responses 1994). The reader is urged to read the whole discussion; it is quite instructive. It was sparked by the letter of A. Jaffe and F. Quinn (1993) entitled 'Theoretical Mathematics: towards a cultural synthesis of mathematics and theoretical physics'. The authors were worried by the local situation in the very active domain of mathematics bordering on mathematical physics. It seemed to them that the standards of physical reasoning (which are considerably lower than those in mathematics) tended to influence unfavourably the standards of today's mathematical research. At the same time they fully recognized the value of cross–fertilization, and suggested some rules of conduct that should be imposed upon all players, in particular the rules of assigning credit. (The word 'theoretical' in the title in the present context is used in a non-standard way, and this usage is not a very happy one because the authors have in mind a mixture of educated speculations, examples, and computer outputs, as opposed to theorems with proud quantifiers.)

> A. When I started as a graduate student at Berkeley, I had trouble imagining how I could 'prove' a new and interesting mathematical theorem. I didn't really understand what a 'proof' was.
>
> By going to seminars, reading papers, and talking to other graduate students, I gradually began to catch on. Within any field, there are certain theorems and certain techniques that are generally known and generally accepted. When you write a paper, you refer to these without proof. You look at other papers in the field, and you see what facts they quote without proof, and what they cite in their bibliography. You learn from other people some idea of their proofs. Then you're free to quote the same theorem and cite the same citations. You don't necessarily have to read the full papers or books that are in your bibliography. Many of the things that are generally known are things for which there may be no known written source. As long as people in the field are comfortable that an idea works, it doesn't need to have a formal written source. W. Thurston, Fields Medal 1983 (Responses 1994, p. 168)

Thurston eloquently argues that the principal goal of the proof is understanding and communication, and that it is most efficiently achieved via personal contacts. His opponents in particular notice that trans-generational contacts can be achieved only via written texts of sufficient level of precision, and that the fate of Italian algebraic geometry should serve as a warning.

> B. We must carefully distinguish between modern papers containing mathematical speculations, and papers published a hundred years ago which we, today, consider defective in rigor, but which were perfectly rigorous according to the standards of the time. Poincaré in his work on Analysis Situs was being as rigorous as he could, and certainly was not consciously speculative. I have seen no evidence that contemporary mathematicians considered it 'reckless' or 'excessively theoretical' [*in the JQ sense, Yu. M.*]. When young Heegard in his 1898 dissertation brashly called the master's attention to subtle mistakes, Poincaré in 1899, calling Heegard's paper 'très remarquable', respectfully admitted his errors and repaired them. In contrast, in his 1912 paper on the Annulus Twist theorem (later proved by Birkhoff), Poincaré apologized for publishing a conjecture, citing age as his excuse. M.W. Hirsch (Responses 1994, p. 187)

> C. Intuition is glorious, but the heaven of mathematics requires much more [...] In theological terms, we are not saved by faith alone but by faith and works [...] Physics has provided mathematics with many fine suggestions and new initiatives, but mathematics does not need to copy the style of experimental physics. Mathematics rests on proof—and proof is eternal. S. Mac Lane (Responses 1994, pp. 190–3)

> D. Philip Anderson describes mathematical rigor as 'irrelevant and impossible'. I would soften the blow by calling it besides the point and usually distracting, even when possible. B. Mandelbrot (Responses 1994, p. 194)

Mandelbrot's contribution is a vehement attack, not only on the abstract notion of rigorous proof, but also on a considerable part of the American mathematical community, 'Charles mathematicians', who allegedly are totalitarian, concentrate on credit assigning, and strive to isolate open minded researchers.

> E. Before 1958 I lived in a mathematical milieu involving essentially Bourbakist people, and even if I was not particularly rigorous, these people—H. Cartan, J.-P. Serre, and H. Whitney (a would-be Bourbakist)—helped me to maintain a fairly acceptable level of rigor. It was only after the Fields medal (1958) that I gave way to my natural tendencies, with the (eventually disastrous) results which followed. Moreover, a few years after that, I became a colleague of Alexander Grothendieck at the IHES, a fact which encouraged me to consider rigor as a very unnecessary quality in mathematical thinking. R. Thom (Responses 1994, p. 203)

Thom's irony requires a slow reading. In what sense did following his natural tendencies have eventually disastrous results? How exactly did becoming a colleague of Grothendieck's influence Thom's thinking? An outsider may remain puzzled whether Grothendieck himself shared Thom's convictions, or whether it was the other way around. Later in the same contribution Thom invokes *rigor mortis* as an appropriate connotation to the idea of mathematical rigour.

> F. I find it difficult to convince students—who are often attracted into mathematics for the same abstract beauty and certainty that brought me here—of the value of the messy, concrete, and specific point of view of possibility and example. In my opinion, more mathematicians stifle for lack of breadth than are mortally stabbed by the opposing sword of rigor. K. Uhlenbeck (Responses 1994, p. 202)

I would like now to summarize, contributing my own share to the general confusion.

First, individually, producing acceptable proofs is an activity that takes arduous training and evokes strong emotional response. A person feels aversion if required to do something contradicting his or her nature. Innate or acquired preference of geometric reasoning or algebraic calculations can inform our career. When we philosophize, we unavoidably rationalize and generalize these basic instincts, and the whole spectrum of our attitudes can be traced back to the feelings of bliss or frustration that overwhelm us during confrontations with intellectual challenges of our *métier.*

Second, socially, we have to rely upon our contemporaries and forebears even when devising a very rigorous proof. Authority in mathematics plays a two–fold role: we acquire from our fathers and peers a value system (what questions are worth asking, what domains are worth developing, what problems are worth solving), and we rely upon the authority of published and accepted proofs and reasonings. Nothing is absolute here, but nothing is less important because of the lack of absoluteness.

Third, epistemologically, all of us who have bothered to think about it know what a rigorous proof is. It has an ideal representation which was worked out by mathematical logicians in this century, but is only more explicit and not fundamentally different from the notion Euclid had. (In this respect, Bourbaki was quite right.) This ideal representation is an imaginary text which step-by-step deduces our theorem from axioms, both axioms and the rules of deduction being made explicit beforehand, say in a version of axiomatic set theory.

If this image arouses in your heart a strong aversion, or at least if you want to be realistic, you may (and should) object that this ideal is utterly unreachable because of the fantastic length of even the simplest formal deductions, and because the closer an exposition is to a formal proof, the more difficult it is to check it. Moreover, since formal deduction strives to be freed of any remnant of meaning (otherwise it is not formal enough), it ends by losing meaning itself.

On the contrary, if this image arouses your enthusiasm, or once again if you want to be realistic, you will agree that the essence of mathematics requires daily maintenance of the current standards of proof. Whether we are engaged in the mathematical support of a vast technological project like a moon-landing, or simply nurture a natural desire to know which assertions have a chance of being true and which do not, we have to resort to the ideal of mathematical proof as an ultimate judge of our efforts.

Even the use of mathematics 'for narrative purposes' as is nicely put by Hirsch is not an exception, because such a narration is built of blocks of solid mathematics to a non-mathematical blueprint.

> An author with a story to tell feels it can be expressed most clearly in mathematical language. In order to tell it coherently without the possibly infinite delay rigor might require, the author introduces certain assumptions, speculations and leaps of faith, for example: 'In order to proceed further we assume the series converges— the random variables are independent—the equilibrium is stable— the determinant is non–zero—.' In such cases it is often irrelevant whether the mathematics can be rigorized, because the author's goal is to persuade the reader of the plausibility or relevance of a certain view about how some real world system behaves. The mathematics is a language filled with subtle and useful metaphors. The validation is to come from experiment—very possibly on a computer. The goal in fact may be to suggest a particular experiment. The result of the narrative will be not new mathematics, but a new description of reality (*real* reality!). M. W. Hirsch (Responses 1994, pp. 186–7)

A beautiful recent example of such a narrative use of mathematics is furnished by D. Mumford's talk at the first European Congress of Mathematicians (Mumford 1992). About mathematical metaphors, see also (Manin 1990).

3 Materials for three case studies

In this section, I present three cases relevant to our discussion: Gödel's proof of the existence of God (1970), the tale of the faulty Pentium chip (1994), and G. Chaitin's claim (1992 and earlier) that a perfectly well and uniformly defined sequence of mathematical questions can have a 'completely random' sequence of answers. For all their differences, these arguments represent human attempts to grapple with infinity by finitary linguistic means, be it the infinity of God, real numbers, or mathematics itself.

Whatever moral lessons (if any) can be drawn from these materials, the reader is free to decide.

Gödel's ontological proof

The third volume of K. Gödel's Collected Works recently published by Oxford University Press contains a note dated 1970. It presents a formal argument purporting to prove the existence of God as an embodiment of all positive properties.

An introductory account by R. M. Adams (Gödel 1995, pp. 388–402) puts this proof into a historical perspective comparing it in particular to Leibniz's argument and discussing its possible place in theoretical theology.

The proof itself is a page of formulas in the language of modal logic (using Necessity and Possibility quantifiers in addition to the usual stuff). It is subdivided into five Axioms and two Theorems. A photocopy of the published version of this page (p. 403) may help the reader.

What does a computer compute, or truth in advertising

In the January 1995 issue of *SIAM News* the front page article 'A Tale of Two Numbers' started with the following lines:

> This is the tale of two numbers, and how they found their way over the Internet to the front pages of the world's newspapers on Thanksgiving Day, embarrassing the world's premier chip manufacturer.

Briefly, it was found that the Intel Corporation's newly launched Pentium chip (the central processing unit in personal computers) contains a bug in its Floating-Point-Divide instruction so that, for example, on calculating

$$r = 4195835 - (4195835/3145727)(3145727)$$

it produces $r = 256$ instead of the correct value $r = 0$.

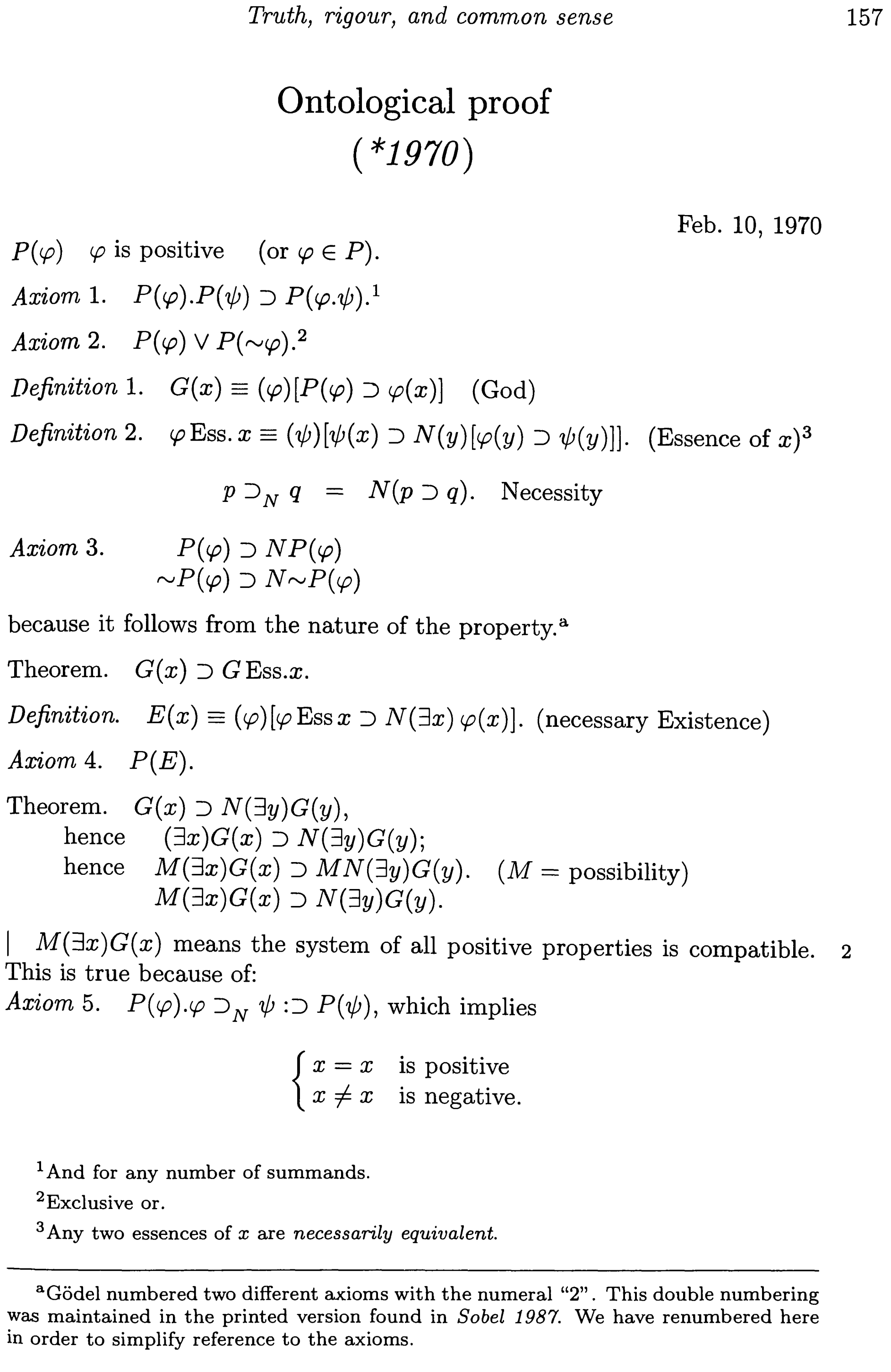
Now, this is not something very unusual. In fact, in all computers the so-called real number arithmetic *is programmed in such a way that it systematically produces incorrect answers* (*round-off errors*). In this particular case a (slightly inflated) public outrage was incited by the fact that in some cases the error was larger than promised (simple-precision when double-precision was advertised).

Completely precise calculations with rational numbers of arbitrary size can be programmed in principle (and are programmed for special purposes). This requires a lot of resources and might need also specialized input–output devices. The ideal Turing machine is highly impractical to implement, and real computers are not designed to facilitate this task.

It is not difficult to imagine a computerized system of decision–making which is unstable with respect to small calculational errors. Stock market or military applications are sensitive to such problems. Here is one more example.

A recent study of sexuality in USA purportedly designed to support epidemiological models of the spread of AIDS did not include the 3 per cent of Americans who do not live in households, i.e. those who live in prisons, in homeless shelters, or on the street. A critic of this study (R. C. Lewontin, *New York Review of Books*, 20 April 1995) reasonably remarks:

> The authors do not discuss it, and they may not even realize it, but mathematical and computer models of the spread of epidemics that take into account real complexities of the problem often turn out, in their predictions, to be extremely sensitive to the quantitative values of the variables. Very small differences in variables can be



Truth, rigour, and common sense 157

Ontological proof
(**1970*)

Feb. 10, 1970

$P(\varphi)$ φ is positive (or $\varphi \in P$).

Axiom 1. $P(\varphi).P(\psi) \supset P(\varphi.\psi)$.[1]

Axiom 2. $P(\varphi) \vee P(\sim\varphi)$.[2]

Definition 1. $G(x) \equiv (\varphi)[P(\varphi) \supset \varphi(x)]$ (God)

Definition 2. $\varphi\,\text{Ess.}\,x \equiv (\psi)[\psi(x) \supset N(y)[\varphi(y) \supset \psi(y)]]$. (Essence of x)[3]

$$p \supset_N q \quad = \quad N(p \supset q). \quad \text{Necessity}$$

Axiom 3. $P(\varphi) \supset NP(\varphi)$
$\sim P(\varphi) \supset N\sim P(\varphi)$

because it follows from the nature of the property.[a]

Theorem. $G(x) \supset G\,\text{Ess.}x$.

Definition. $E(x) \equiv (\varphi)[\varphi\,\text{Ess}\,x \supset N(\exists x)\,\varphi(x)]$. (necessary Existence)

Axiom 4. $P(E)$.

Theorem. $G(x) \supset N(\exists y)G(y)$,
hence $(\exists x)G(x) \supset N(\exists y)G(y)$;
hence $M(\exists x)G(x) \supset MN(\exists y)G(y)$. ($M$ = possibility)
$M(\exists x)G(x) \supset N(\exists y)G(y)$.

| $M(\exists x)G(x)$ means the system of all positive properties is compatible. 2
This is true because of:
Axiom 5. $P(\varphi).\varphi \supset_N \psi :\supset P(\psi)$, which implies

$$\begin{cases} x = x & \text{is positive} \\ x \neq x & \text{is negative.} \end{cases}$$

[1] And for any number of summands.

[2] Exclusive or.

[3] Any two essences of x are *necessarily equivalent.*

[a] Gödel numbered two different axioms with the numeral "2". This double numbering was maintained in the printed version found in *Sobel 1987*. We have renumbered here in order to simplify reference to the axioms.

FIG. 1. Gödel's ontological proof. From Kurt Gödel, *Collected Works, Volume III: Unpublished Essays and Lectures* (1995), reproduced by permission of Oxford University Press.

> the critical determinant of whether an epidemic dies out or spreads catastrophically, so the use of inaccurate study in planning countermeasures can do more harm than does total ignorance.

The problem of understanding what is computed by a computer becomes also more and more relevant with the spread of computer-assisted proofs of mathematical theorems. I quote M. Hirsch once again (Responses 1994, p. 188):

> Oscar Lanford pointed out that in order to justify a computer calculation as a part of proof (as he did in the first proof of the Feigenbaum cascade conjecture), you must not only prove that the program is correct (and how often is this done?) but you must understand how the computer rounds numbers, and how the operating system functions, including how the time-sharing system works.

Randomness of mathematical truth

Following A. N. Kolmogorov's, R. Solomonoff's, and G. Chaitin's discovery of the notion of complexity and a new definition of randomness based upon it, Chaitin constructed an example of an exponential Diophantine equation

$$F(t; x_1, \ldots, x_n) = 0$$

with the following property (Chaitin 1992). Put $\varepsilon(t_0) = 0$ (respectively, 1), if this equation has, for $t = t_0$, only finitely (respectively, infinitely) many solutions in positive integers x_i. *Then the sequence* $\varepsilon(1), \varepsilon(2), \varepsilon(3), \ldots$ *is random.* (Chaitin in fact has written a program producing F. The output is a 200-page long equation with about 17 000 unknowns).

This is a really subtle mathematical construction, using among other tools the Davis–Putnam–Robinson–Matijacevič presentation of recursively enumerable sets. The epistemologically important point is the discovery that randomness can be defined without any recourse to physical reality (the definition is then justified by checking that all the standard properties of 'physical' randomness are present) in such a way that the necessity to make an infinite search to solve a parametric series of problems leads to the technically random answers.

Some people find it difficult to imagine that a rigidly determined discipline like elementary arithmetic may produce such phenomena. Notice that what is called 'chaos' Mandelbrot-style is a considerably less sophisticated model of random behaviour.

Bibliography

Chaitin, G. (1992). *Information, randomness and incompleteness. Papers on algorithmic information theory.* World Scientific, Singapore.

Gödel, K. (1995). Ontological proof. In *Kurt Gödel: collected works*, Vol. 3, (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Soloway, and J. van Heijenoort), p. 403. Oxford University Press, New York.

Hildebrandt, S. (1995). *Wahrheit und Wert mathematischer Erkenntnis.* Carl Friedrich von Siemens Stiftung, München.

Jaffe, A. and Quinn, F. (1993). Theoretical mathematics: toward a cultural synthesis of mathematics and theoretical physics. *Bull. American Math. Soc.*, **29**, 1–13.

Manin, Yu (1990). Mathematics as metaphor. In *Proceedings of the International Congress of Mathematicians, Kyoto,* Vol. 2, Mathematical Society of Japan and Springer-Verlag, pp. 1665–71.

Mumford, D. (1994). Pattern theory: a unifying perspective. *First European Congress of Mathematics* (Paris 1992), Vol. 1, pp. 187–224. Birkhäuser, Basel.

Responses (1994). Responses to 'Theoretical mathematics etc.', by A. Jaffe and F. Quinn. *Bull. American Math. Soc.*, **30**, 161–77.

Weil, A. (1980). *Collected Papers, Vol. 2*, Springer-Verlag, Berlin.

Max-Planck-Institut für Mathematik
Gottfried-Claren-Strasse 26
D-53225 Bonn
Germany
email: manin@mpim-bonn.mpg.de

9

How to be a naturalist about mathematics

Penelope Maddy

'Naturalism', as I use the term, is a metaphilosophical principle describing the proper relations between philosophy and methodology, between philosophical theorizing about a given practice and methodological decisions about how that practice should be pursued. This usage springs from work of Quine and Putnam in the philosophy of science, but I have recently applied it to mathematics in ways that depart substantially from the views of these authors.[1]

My focus so far has been on how naturalism affects the evaluation of set-theoretic methodology, that is, on its implications for particular, practical decisions. But a principle on the relations between philosophy and methodology can be expected to raise philosophical issues as well, and perhaps to have philosophical ramifications. I will focus here on the philosophical aspects of naturalism. This is a big topic, so I can only attempt a preliminary sketch of the terrain.

1 Naturalism in mathematics

Let me begin with a review of the type of methodological quandary that motivates the move to naturalism in the first place; I have in mind the well-known methodological difficulties raised by contemporary set theory. The most familiar of these centers on the continuum hypothesis (CH): should the work of Gödel and Cohen be regarded as settling the continuum problem, or does a mathematical question remain, amenable to solution by mathematical methods? The difficulty this question raises is a methodological one because it concerns the proper pursuit of set theoretic mathematics: should the CH continue to be pursued as an open problem? Less familiar examples of the same form arise in descriptive set theory, concerning, for example, the Lebesgue measurability of definable sets of reals, the existence of perfect subsets of uncountable definable sets of reals, and similar questions.[2]

A methodological difficulty of a slightly different sort involves the axioms of set theory: what justification can properly be offered for an axiom candidate? This problem becomes particularly dramatic if we assume, in answer to the first difficulty, that the CH and the questions of descriptive set theory are mathematically legitimate, and if we undertake a search for new set-theoretic axioms to

The support of the NSF (SBR-9320220) is gratefully acknowledged. My thanks go to Jeffrey Barrett, Mark Balaguer, and Alexander George for helpful comments on earlier drafts.

settle them. But it arises just as inevitably for the familiar axioms of ZFC. On what grounds do we justify the adoption of these axioms? [3]

Even the most superficial survey of the literature on these methodological difficulties reveals that the debates often turn philosophical, that is, that they often involve discussions of traditional philosophical issues. For example, we hear such familiar ontological questions as these: do mathematical objects or concepts exist? Are they like fictional objects? Is there no mathematical ontology at all? Is mathematics merely a matter of determining which conclusions follow logically from which premises? Other examples are more generally metaphysical: is mathematical existence objective? Do we create mathematical things by acts of mind or definition? Are mathematical objects located in space–time? Are they abstract? Acausal? And so on.

These questions and their purported answers produce a bewildering array of philosophical stances on the methodological difficulties of set theory. Consider, for example, a Simple Realism, which holds that set theory is the study of an objectively existing world of sets. From this point of view, CH and the rest are still legitimate mathematical questions, with determinate answers, and to justify an axiom candidate is to provide evidence that it is true in the world of sets. Some versions of Simple Realism even posit a strong analogy between mathematics and natural science, so that the implication of lower level truths counts as evidence for the truth of more theoretical statements or axiom candidates.[4]

But there are other forms of Realism, which hold just as staunchly that mathematical things exist objectively, but which nevertheless provide different answers to our difficulties. For example, consider Plentiful Platonism, the view that there exists an objective world of sets corresponding to each and every consistent theory in a first-order language with $\in$ as its sole non-logical symbol.[5] On such a view, CH has no determinate truth value—it is true of some sets, false of others—and the same goes for the open questions of descriptive set theory. Furthermore, any relatively consistent axiom candidate is on a metaphysical par with any other, because each and every consistent theory corresponds to its own objective world of sets.

Outcomes intermediate between these two Realisms might result in the view that (non-plentiful) mathematical things exist objectively, but that they are somehow incomplete. On appropriate accounts of this incompleteness, questions like CH would have no answers, but the requirement that axioms be true would nevertheless rule out some relatively consistent candidates. One might also consider views that countenance mathematical concepts, rather than objects. Versions of Conceptualism parallel to these three forms of Realism would hold that CH is either true or false to the concept of set (analogous to Simple Realism), or that there is a concept of set corresponding to every consistent set theory (analogous to Plentiful Platonism), or that the concept of set is vague in spots, so that CH has no determinate truth value (analogous to Incomplete Object Realism). As the methodological consequences of these versions of Conceptualism will also run parallel to those of their corresponding Realisms, I will not treat them separately in what follows.

Metaphysical questions about the nature of the Realist's objects—are they spatio-temporal? Acausal? etc.—often appear in the course of philosophical arguments against Realism, especially Simple Realism. In the most familiar of these, the qualities usually taken as characterizing the abstractness of mathematical things—their non-spatiotemporality and acausality—are argued to preclude a satisfactory epistemological theory of our knowledge of them.[6] These and other worries often motivate the search for non-Realistic philosophical accounts of mathematics. One familiar example is the view that mathematics is not a study of objects (or concepts), but a purely logical inquiry into which conclusions follow from which hypotheses. From this perspective, the only legitimate questions in the vicinity of the CH are of the sort that Gödel and Cohen, Levy and Solovay,[7] have settled. The simplest version of this position—which I will call 'Glib Formalism'[8]—would go on to hold, in considerable agreement with Glib versions of Plentiful Platonism,[9] 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.

Though there are many others, I will mention only one more popular philosophical position on mathematics, namely, Fictionalism: the view that mathematical theories are like fictional stories, that mathematical objects are metaphysically akin to the fictional characters of imaginative literature.[10] The precise status of mathematical entities will then follow from the accompanying theory of fictional objects, on which philosophers have a wide range of views. As far as our methodological questions are concerned, the Fictionalist might hold that the CH is a pseudo-question, comparable to the question, 'how long is Hamlet's nose?'; the answer is neither explicit in, nor derivable from, the story that defines the character, so it has no answer. Like the Glib Formalist and the Glib Plentiful Platonist, a Glib Fictionalist might say that all (consistent) mathematical stories are on a par, that there is no principled way to justify a choice between them.

Of course, this is just the barest beginning; the range and the subtleties of this philosophical literature, of these ontological and metaphysical debates, goes far beyond anything that has been so much as hinted at here. Fortunately, only two relatively simple observations are crucial for our purposes. The first is that the introduction of these philosophical controversies into the discussion of proper mathematical method brings with it a wealth of forms and styles of argument that were not previously to be found in the mathematical practice itself. To take just a few examples, there are pro-realistic arguments in terms of a purported analogy between mathematics and science; there are anti-realistic arguments based on theories in epistemology or cognitive science or on physicalistic intuitions; there are ontological arguments based on the role of mathematics in applications, debates over the nature of applied mathematics, debates over whether or not science could be conducted without mathematics; there are metaphysical debates about the nature of truth and objectivity, and much more. My point is just that these are not the sorts of considerations that ordinarily play a role in the practice of modern mathematics. New evidential standards are being introduced along with the philosophy.

Furthermore, the logic of this literature and these arguments suggests that the methodological decisions of contemporary set theory cannot be reached until (at least some of) these outstanding philosophical disputes have been resolved. Not only are new standards introduced, but these new standards must be exercised before the mathematical work can carry on with confidence. My second observation is just this: given that the most fundamental philosophical controversies have not been, and probably never will be, put to rest, this state of dependency is a most disheartening one for the mathematician out to make practical decisions on how to proceed.

In fact, a look at the historical record reveals that situations of this structure have arisen before in modern mathematics, that is, situations in which philosophical argumentation has entered a debate about mathematical method. I have in mind two such debates, one over whether or not impredicative definitions should be allowed, the other over whether or not the Axiom of Choice should be adopted. In both these cases, philosophical considerations of the sort rehearsed above were raised and contested, and in both cases, the fate of the method seemed to hinge on the outcome of those philosophical contests. But, again in both cases, this is not how things turned out. Impredicative definitions and the Axiom of Choice are now respected tools in the practice of contemporary mathematics, while the philosophical issues remain subjects of ongoing controversy. The methodological decision seems to have been motivated, not by philosophical argumentation, but by consideration of what might be called, for want of a better expression, mathematical fruitfulness (for example, the classical theory of reals, among other things, in the case of impredicative definitions, and fundamental results in a mind-boggling array of fields in the case of choice).

What are we to make of this? One response would be to insist that the mathematical community has been too hasty in its embrace of these disputed methods, that they are being used, as it were, without justification, without their essential philosophical underpinnings. The response I propose turns this position on its head: given that the methods are justified, that justification must not, after all, depend on the philosophy. Mathematical naturalism, as I understand it, is just a generalization of this conclusion, namely, that mathematical methodology is properly assessed and evaluated, defended or criticized, on mathematical, not philosophical (or any other extra-mathematical) grounds. The particular instances of mathematical fruitfulness that played the decisive roles in the impredicativity and choice controversies stand as ready examples of the type of 'mathematical grounds' that I have in mind.

This use of the term 'naturalism' derives from the scientific naturalism of Quine and Putnam:

> ... naturalism ... sees natural science as an inquiry into reality, fallible and corrigible but not answerable to any supra-scientific tribunal, and not in need of any justification beyond observation and

> the hypothetico-deductive method. (Quine 1975, p. 72)
>
> ... it is silly to agree that a reason for believing that p warrants accepting p in all scientific circumstances, and then to add 'but even so it is not *good enough*'! Such a judgement could only be made if one accepted a trans-scientific method as superior to the scientific method; but this philosopher, at least, has no interest in doing *that.* (Putnam 1972, p. 356)

However, neither of these scientific naturalists is so respectful of the methods of mathematics; they both see the ultimate justification of mathematical practice to lie (via a philosophical argument) in its applications in science, not in the usual justificatory apparatus of mathematics itself. This approach leaves out large parts of contemporary mathematics (the as yet unapplied part), and submits the rest to an extra-mathematical tribunal. My suggestion is that the success story of modern mathematics has been won by the application of actual mathematical methods, not by exercise of the extra-mathematical standards Quine and Putnam propose, and thus, that the methodologist should judge those methods on their own terms.

The picture, then, is this. All naturalists begin their study within natural science; this is scientific naturalism. All scientific naturalists notice that mathematicians employ methods different from those of natural scientists. The response of the Quinean or science-only naturalist is to regard mathematical claims as justified only in so far as they are supported by scientific, as opposed to mathematical, methods. In contrast, the response of the mathematical naturalist—influenced by the observation that mathematics has flourished by its own methods, not by those recommended by the science-only naturalist—opts to evaluate mathematical methods in their own terms, opts not to hold mathematical methods answerable to natural science.

This is how I will understand the proper naturalistic approach to the study of mathematical method. Our next question is: how is it to be carried out?

2 The naturalistic methodologist

To return to our original motivation, the naturalistic methodologist hopes to discover what constitutes good grounds, that is, rational grounds, for settling the methodological difficulties of contemporary set theory, for example, for deciding whether or not CH and the open questions of descriptive set theory remain legitimate mathematical concerns, or for defending or criticizing potential new axioms. As a naturalist about mathematics as well as science, this same naturalistic methodologist proposes to investigate these methodological difficulties using standards drawn from mathematics itself, not from any extra-mathematical source.

At first glance, it might seem it would be easy to focus attention exclusively on intra-mathematical considerations; it might seem that this could be done simply by focusing on what mathematicians actually say and do. But we've already seen that actual mathematical discourse includes philosophical discussion, for example, when Hadamard in 1905 argues for the Axiom of Choice, (see (Hadamard *et al.* 1982)), or when Gödel (1947) argues for the legitimacy of the continuum problem. How, then, is the naturalist to distinguish the properly mathematical from the extra-mathematical incursions?

One approach would be to expound and defend a principled distinction between mathematics and philosophy, or more generally, between mathematics and everything else. I doubt this is possible; at any rate, it is not the course I will take here. Instead, following the lead of our historical examples, I will propose a naturalized model of practice. This naturalized model will not include considerations that seem, by extrapolation from the historical cases, to be methodologically irrelevant, and it will include more detailed analysis and development of the considerations that remain.[11] My claim is that this purified and amplified model provides an accurate picture of the actual justificatory structures of contemporary set theory and that this justificatory structure is fully rational. I will say more below about how this claim can be tested, about how the model can be assessed for adequacy.

Before I begin, let me note that the philosophical incursions I will be eliminating in the naturalistic model do play an important and legitimate role in practice; the naturalist's claim is only that they are not justificatory. Their actual role, I have claimed elsewhere,[12] is inspirational: the mathematician may be inspired to certain investigations by her extra-mathematical beliefs, though the proper defense of that work lies solely in the mathematics it produces. It would be silly not to admit the importance and interest of such inspirations, but as our focus here is on the justification of set-theoretic methods, inspirations will be set to one side.[13]

What, then, is the counsel of our historical cases? The negative counsel is that certain, typically philosophical issues are ultimately irrelevant to the defense or criticism of mathematical methods. Among these are issues about the metaphysical status of mathematical things: are they objective? Are they spatio-temporal? And so on. To apply this negative counsel to the cases that interest us will require us to extrapolate from the historical cases by identifying elements and themes in contemporary discussions that seem analogous to those historically irrelevant elements.

The historical cases also provide positive counsel. There is a pattern in what remains after the extraneous is eliminated, a pattern in the considerations that *are* relevant, in the considerations that *are* ultimately decisive. In both our cases, the community eventually reached a consensus that the controversial method was admissible because it led to certain varieties of mathematics, that is, because it was an effective means to particular desirable ends. Thus the positive counsel of history is to frame a defense or critique of a particular method in two parts: first, identify a goal (or goals) of the relevant practice, and second, argue that the method in question either is or is not an effective means towards that goal. In detail, we should expect that some goals will take the shape of means towards higher goals, and that goals at various levels will conflict, requiring a subtle assessment of weights and balances. But the simple counsel remains: identify the goals and evaluate methods by their relations to those goals.

Following these positive and negative counsels from history, the naturalist eliminates the methodologically irrelevant distractions from the actual practice and brings out the means/ends considerations implicit there, all with the aim of drawing out sound methodological arguments. I have attempted to apply this technique to aspects of our set-theoretic difficulties in a number of places.[14] As our focus here is on the philosophical aspects of naturalism, rather than its practical repercussions, I will confine myself to sketching a few illustrations for future reference.

For a sense of how the naturalist's technique plays out in the case of CH, consider Gödel's famous paper, 'What is Cantor's continuum problem?'. Gödel argues that the CH remains a legitimate mathematical question despite its independence,[15] and his most conspicuous argument is decidedly metaphysical:

> ... the set-theoretical concepts and theorems describe some well-determined reality, in which Cantor's conjecture must be either true or false. (Gödel 1947; p. 260 of 1990 version)

This ontological claim is defended by philosophical means, in terms of an analogy between mathematics and science, intuition and sense perception, and so on. But when the naturalist stubbornly sets that material aside, a second line of argument emerges:

> ... the question of the objective existence of the objects of mathematical intuition ... is not decisive for the problem under discussion here [i.e., the legitimacy of CH]. (Gödel 1964, p. 268)

> ... it is possible to point out ways in which the decision of a question, which is undecidable from the usual axioms, might nevertheless be obtained. (Gödel 1947; p. 260 of 1990 version)

This last claim is defended mathematically, by pointing to mathematically reasonable ways of extending the axioms of ZFC. This is the hint that the naturalist takes up; in so far as a problem is amenable to mathematically sound solution, it is obviously, *ipso facto*, a mathematically legitimate question.

So, the naturalist asks, how strong is this argument for the legitimacy of the CH? In other words, how plausible is the claim that there is a mathematically sound solution, that there are mathematically defensible new axioms that will settle it? As it happens, there is more room for this type of confidence in the legitimacy of the open questions of descriptive set theory. Set theorists already have two different axiom candidates that settle those questions ($V = L$ and LC),[16] and they settle them in different ways. Here the (mathematical) case for the legitimacy of the questions is at least as strong as the (mathematical) case for one or the other of the new axiom candidates.

There is a strong consensus among set theorists against one of these axiom candidates ($V = L$) and for its competitor (LC). I have attempted to give naturalistic arguments, in terms of goals and effective means, for the conclusion that this consensus is rationally justified.[17] But even if this line of thought is cogent, the situation for CH remains problematic. Neither the goals identified in these arguments nor the general methods justified in terms of them seem to provide any guidance for the case of CH, and a survey of the literature for other clues is not encouraging. There is no general consensus on CH, and the methods on which there is consensus are provably inadequate for its solution.

Under the circumstances, some contemporary set theorists say things like 'the CH may be neither true nor false', and such statements are customarily explicated in philosophical terms. In opposition to the Simple Realist, the Plentiful Platonist says that a robust correspondence obtains between CH and some set-theoretic universes and between not-CH and other set-theoretic universes, so that CH has no determinate truth value. Or the Formalist says that there are not any such set-theoretic universes, that 'CH is true' or 'CH is false' can only mean 'CH follows from our current theory of sets' or 'not-CH follows from our current theory of sets', and that neither of these is true. Or the Fictionalist says much the same, with 'our current theory of sets' replaced by some story about sets.

I think it takes very little extrapolation from the historical record to classify this sort of discussion as philosophical and extra-mathematical. And again, if this material is eliminated, the naturalist finds a simple surrogate for the set theorist's remark, a surrogate that bypasses the philosophically charged word 'true'. In brief, the idea is simply to replace 'CH may be neither true nor false' by 'there may be no mathematically sound grounds on which to settle CH one way or the other' in the naturalist's model. More fully, the proposed argument for the legitimacy of the open questions of descriptive set theory defends two set-theoretic methods that potentially conflict, methods which I have called Maximize and Unify. So far, it has been possible to satisfy both at once, but in the case of CH, this might not be possible. If so, Maximize might reasonably be taken to overrule Unify, which might lead set theorists to adopt a variety of set theories, all with different values for the size of the continuum.[18] This possibility is the naturalist's de-philosophized version of the discouraged set theorist's remark.

In fact, I think most seemingly robust uses of 'true' in set theoretic discussion will disappear in the naturalist's model. Things like 'Borel determinacy turned out to be true' can be naturalized as the claim that Borel determinacy turned out to be a theorem of ZFC. A statement like 'Since the Axiom of Choice is true, the full axiom of determinacy must be false' can be replaced by the observation that set theory with the Axiom of Choice serves the ends of set theory better than set theory without it, so that adding the Axiom of Determinacy to our preferred theory is not an attractive option (assuming Consistency is an overarching goal). The proposal that everything true in V is actually true in some V_κ takes the naturalistic form: if you can argue for φ, conclude that there is a κ such that V_κ thinks φ. In discussions of axiom candidates, a statement like 'Such-and-such an axiom is probably true' can be replaced by 'Such-and-such an axiom is probably an effective method for achieving the goals of set theory, probably better than the alternatives'. On this reading, the evidence for the statement of 'probable truth' will come in terms of what would follow if the axiom candidate were adopted. And so on.

Some stubborn hold-outs will remain, of course. For example, any claim of the form 'if axiom candidate T is true, ...' that cannot be replaced by something of the form 'if axiom candidate T were added to the standard list of axioms, ... ', any truth claim that is viewed as independent of what set theory is adopted and what consequences that set theory has—any such claim would involve a more robust notion of 'truth'. Here the guidance of history again prompts the naturalist to excise such considerations as extra-mathematical.[19]

Suppose, then, that the naturalist has extrapolated from the historical record and proposed a naturalized model of set theoretic practice. In this model, some elements of the actual practice are absent, namely, those elements that the historical record suggests are actually extra-mathematical. In what remains, the purely mathematical elements are highlighted, and various methodological choices are defended by arguing that mathematics, in general, and set theory, in particular, have various internal goals, and that the methods defended (criticized) are the most effective available[20] means (ineffective means) for achieving those goals. The result—this purified and amplified version of actual practice—claims to reflect the underlying justificatory structure of that actual practice and to show that this structure is rational. In other words, the naturalist claims:

(a) that the identified goals are actual goals of mathematics, in general, and set theory, in particular (and, in complex cases, that they interact as they are portrayed to do);

(b) that the methods defended (criticized) really are the most effective available means (ineffective means) for achieving those goals; and

(c) that there are no other available considerations, overlooked or improperly excised, that in fact support or undermine the methodological conclusions drawn. How are these claims to be tested?

I think the naturalist's model can be assessed in at least three ways. First, we can ask whether it yields accurate readings of historical cases that have, in fact, been resolved. As the naturalist's techniques are derived from our two historical cases, some success on this first test is built in. Second, the naturalist's methodological arguments can be tested for their plausibility in the eyes of contemporary practitioners. On this test, the jury is out. The conclusions argued for (for example, that $V = L$ should be rejected) reflect the general consensus of the community, but this by no means guarantees that the arguments provided in support of that consensus will be judged to be the correct ones. Third, and finally, the naturalist's arguments can be viewed as predictions—that these debates will eventually be resolved in these ways on these grounds—and eventually history will report on the fate of these predictions. So the naturalist's claims are eminently falsifiable.

The issues raised by these questions of testing intertwine with a range of questions standardly put to naturalistic positions, beginning with the scientific naturalism of Quine and Putnam. For example, naturalists are asked: is the word of the practitioner to be taken as gospel? Is naturalism purely descriptive?

Is the naturalistic study of methodology just sociology of science? I would answer all these questions in the negative, and I think some light will be cast if I say a bit more about why this is so.

First, consider part (b) of the naturalist's claim: that the methods defended (criticized) are the most effective means (ineffective means) for achieving the identified goals. This is an objective matter, about which individual practitioners, and even the entire community, could be mistaken. Of course, the conclusion that the entire community is mistaken on such a matter should be viewed with considerable skepticism; more likely the naturalistic observer is mistaken about the means being used or the ends being pursued. But it remains logically possible that the community could falsely believe that method A rather than method B is best suited to its goals. So, on this score, in principle, expert testimony is controvertible.

Now consider part (a): that the identified goals are the actual goals of the practice. Notice again that the individual practitioner or the community might reject a sound naturalistic evaluation of one of its methods for failure of self-analysis, for failure to recognize the goal identified in the argument as its own. Again, this logical possibility is an unlikely one, to be viewed with considerable skepticism, but it reinforces the conclusion that expert testimony can conceivably be defeated by other considerations. We should also note that this (remote) possibility of error only concerns the identification of goals, not the choice of goals. The naturalist has no independent grounds on which to defend or criticize the actual goals of the practice.

Finally, in connection with part (c)—the claim that no relevant justificatory considerations have been overlooked or improperly excised—it must be noted that the naturalist departs from absolute faith in the testimony of practitioners at the very outset, by insisting that extra-mathematical considerations are irrelevant to sound evaluation of mathematical methods and by undertaking to eliminate them. According to the naturalist, the individual practitioner can be wrong about the evaluation of certain methods, for example, by founding them on philosophical views that inspire rather than support their use.

In fact, naturalistically unacceptable justifications of this sort are often idiosyncratic or special to a certain segment of the community, in which case the naturalist is only eliminating what fails of stable consensus. A more difficult question is whether or not the entire community could rightly be said to err in its evaluation of a particular method by basing it on external, thus naturalistically unacceptable, grounds. If those grounds are 'sociological'—for example, intellectual fashion, browbeating from philosophers, political pressure, etc.—and if the efforts of naturalistic philosophers of science and mathematics are successful in drawing a principled distinction between the sociological and the scientific, the sociological and the mathematical,[21] then I suppose it could be rightly said that the entire community is in error. But I have offered no principled distinction between philosophical and mathematical considerations; I have used historical cases and our rough-and-ready sense of the distinction to extrapolate from those historical cases. If the entire community were to agree on the force of a certain consideration that I label as extra-mathematical philosophy, if that consideration ultimately led the community to a stable methodological decision, this would count as reason to reject my proposal on where the line between extra- and intra-mathematical should be drawn, not as a reason to count the entire community as mistaken.

In sum, then, practitioners, like anyone else, can be wrong about facts: they can be wrong (though this is unlikely) about what methods are effective for furthering a certain goal; they can be wrong (though this is again unlikely) about what their goals actually are; and they can be wrong (this point has been our focus) about justification, for example, by mistaking inspiration for justification. Conversely, they cannot be wrong, collectively, in adopting a certain goal, and they cannot be wrong, collectively, in making a methodological decision on certain grounds.[22] It follows, in answer to the question for naturalism, that the testimony of practitioners should not be taken as gospel. But I should emphasize that the question concerns the logical possibility of error; it is a conceptual question. In naturalistic practice, I suggest that the first two possibilities of error can safely be ignored, that the opinions of practitioners can safely be taken as the central guides to the formulation of naturalistic arguments for or against particular methods, and that the reactions of practitioners can safely be used as central tests in evaluating those arguments. The only likely possibility of error, the one which the naturalist must guard against, is that extra-mathematical considerations will confuse and distort methodological discussions.

If the word of practitioners is not to be taken as gospel, it might seem that the naturalist must depart from pure description, and hence from sociology, but in fact this is not so clear. The first stage of the naturalist's analysis—that of constructing the purified model of practice—can be viewed as a more-or-less sociological undertaking: guided by the structure and outcomes of various historical disputes, the naturalist attempts to prune away contemporary considerations that seem likely to prove irrelevant in the long run (for example, certain typically philosophical arguments), and to highlight and enhance contemporary considerations that seem likely to be decisive (for example, various means/ends analyses). Granted, the practitioner's pronouncements are not taken at face value, but working sociologists certainly admit that the testimony of subjects is sometimes less that fully trustworthy, that an accurate description of a practice will sometimes require discounting the practitioner's reports.

It is only at the second stage—when arguments are offered within the practice for or against particular methodological choices—that the naturalist departs from sociology, and, for that matter, from natural science itself. At this point, the naturalist is using the methods of mathematics, not those of science, and she is doing so exactly as a mathematician might do, except that her choices among the available styles of argument are guided by the results of the first stage of her analysis. In other words, in this second stage, the naturalist is functioning within mathematics, just as a mathematician might, except that she uses only those styles of argument that her previous analysis suggests are the effective ones, the instrumentally rational ones. But the arguments themselves show no sign of sociology; the naturalist does not argue 'this method is preferable because it conforms to previous practice', but 'this method is preferable because it is the most effective method available for achieving this goal'. At this stage, the naturalist is doing what the sociologist might call 'going native'. Going native is something the sociologist avoids as detrimental to scientific objectivity, but for the naturalistic methodologist, it marks one desired payoff of the analysis.

The third and final stage of the naturalist's analysis involves the sort of testing sketched earlier: the arguments produced in the second stage, under the guidance of the first stage analysis, are put to the tribunal of practitioners and of history. This process is also carried out within natural science. A negative outcome would cast doubt on the accuracy of the first stage analysis or on the quality of the second stage arguments, perhaps both.

In sum, then, the naturalistic methodologist differs from the sociologist in going native, or at least, in the motivation for going native: while the sociologist might participate in activities like the naturalist's second stage argumentation—for example, in order to test the first stage analysis—the naturalist has the added (admittedly faint) hope of clarifying the justificatory structure of the practice for the practitioners themselves, and thus, of contributing to that practice. In addition, the desired outcome of the naturalist's third stage is an evaluation of the rationality of certain methodological decisions, a style of conclusion the sociologist is unlikely to draw. For this, what the naturalist needs—in addition to 'mere' sociological observations—is a generic notion of instrumental reasoning as a legitimate variety of practical reasoning. Given that rather modest tool, the naturalist can judge effective means/ends arguments to be fully rational.[23]

I hope this somewhat meandering discussion provides some insight into how a naturalistic evaluation of set-theoretic methodology might proceed and how its results could be evaluated. Let us turn now from methodology to philosophy.

3 The naturalistic philosopher

Notice that this proposed naturalism about mathematics differs markedly from naturalism about science in its treatment of philosophical considerations. Consider, for example, the typically metaphysical claim that mathematical objects exist objectively and non-spatiotemporally. The mathematical naturalist holds these issues to be external to mathematics proper and thus irrelevant to methodological decision-making, but the analogous claim about physical objects—that they exist objectively and spatiotemporally—is part and parcel of scientific thinking. This is not true of the corresponding mathematical questions, which is why the mathematical naturalist undertakes to eliminate them from methodological arguments.

For these reasons, the mathematical naturalist pursuing questions of methodology ignores traditional philosophical questions such as 'are mathematical things objective or subjective?', 'is their existence dependent on our theories or definitions?', 'are mathematical objects incomplete?', 'are they more like fictional objects or physical objects', 'are the axioms true in the real world of sets?', and so on. But perhaps methodology is not the only subject of interest. From a wider perspective, what does the mathematical naturalist say about metaphysical questions like these? Are they pseudo-questions? Or is the naturalist required to hold that there are no mathematical objects? Or that mathematical statements are not true? *Qua* methodologist, the naturalist has nothing to say on these topics, but *qua* philosopher?

The relevant notion of pseudo-question arises from Carnap's distinction between internal and external questions. Within the linguistic framework of mathematics, the internal question—are there numbers?—gets a quick and easy answer: yes. But philosophers, on Carnap's reading, often want to ask another question, a question external to the mathematical linguistic framework, a question that precedes and motivates the adoption of that framework; this question is also phrased—are there numbers?—but it has no easy answer. Carnap argues that the only legitimate external question is a pragmatic one—'are there good reasons to adopt the linguistic framework of number talk?'—and that the philosopher's external question, asked without the support of a linguistic framework to provide the grounds for answering it, is in fact a pseudo-question.

Carnap took the same view of philosopher's questions such as 'is there an external world?' or 'are there *really* atoms?'. Quine, the scientific naturalist, replied that these questions are in fact internal to science, questions to be answered by the very scientific methods that answer Carnap's internal questions. Quine does this by assimilating Carnap's 'pragmatic' considerations to ordinary scientific considerations. In so far as metaphysicians insist on the extra-scientific status of their questions, those questions will be immune to Quine's move and remain pseudo-questions, but short of this, many traditional debates can be brought within the range of naturalistic philosophy in this way.

But we have seen that mathematics does not run parallel to science in this respect; mathematics itself takes no stand on the status of its entities. As only mathematical considerations are relevant to mathematical methodology, it follows that metaphysics is methodologically irrelevant. But if the mathematical naturalist is also a scientific naturalist,[24] as I have been assuming, then there is still room for a naturalistic philosophy of mathematics, for a scientific study of the practice of mathematics. This study would reasonably undertake questions like: what is the relationship between mathematics and science? Is the language of pure mathematics best accounted for as analogous to scientific language or to fictional language? Do mathematical things exist in the same sense as physical ones? The mathematical naturalist sets these metaphysical questions aside for purposes of assessing mathematical methods, but from the scientific naturalist's point of view, they are legitimate questions, not pseudo-questions, regardless of their irrelevance to methodology.

Scientific naturalists have always held that within science there is room for a scientific study of science, a study of scientific language, scientific truth, scientific method. What I am suggesting is that there is also room for a parallel and even comparative study, within science, of mathematical language, mathematical truth, mathematical method. The difference is that the scientific study of science uses the same general methods and pursues the same general goals as its object, while the scientific study of mathematics uses scientific, not mathematical methods, and pursues scientific, not mathematical goals. It is conceivable, if only barely, that a crackerjack scientific study of scientific language or scientific truth might properly influence the practice of science itself, given that the study and the practice share the same methods and goals. But the mathematical naturalist notes that mathematical methods and goals are not those of science, that the scientific study of mathematical method is extra-mathematical and, for that reason, cannot properly interfere with mathematical practice.

Thus, mathematical naturalism has consequences not only for our practical evaluation of mathematical methods, but also for our naturalized philosophical study of mathematics: that study must leave the practice, or rather, the naturalistic methodologist's purified and amplified model of the practice, untouched. Consider, for example, the effects of our naturalistic methodologist's account of the status of CH. If a version of Simple Realism insists that there is a fact of the matter about CH and that the set theorist's job is to find an axiom system that settles it, then the possibility that it might one day be rational to sacrifice Unify to Maximize is ruled out; from the mathematical naturalist's point of view, this is an unacceptable interference of metaphysics in methodology. Similarly, a Glib Formalism or a Glib Plentiful Platonism that denies there can be rational mathematical reasons for choosing one consistent axiom system over another also conflicts with the naturalist's reading of practice.

On the other hand, a Subtle Formalist could hold that mathematics is the study of what follows from which hypotheses, that all consistent axiom systems are metaphysically on a par, but that there are rational reasons, springing from the goals of mathematics itself, that justify the choice of one axiom system over another for extensive study. This Subtle Formalist's purely methodological inquiries might well coincide with the work of the naturalistic methodologist. The same goes for a suitably Subtle Plentiful Platonism: every consistent axiom system correctly describes some world of sets, but there might be sound mathematical reasons for preferring to study one world rather than another. Similar Subtle versions of Fictionalism can also be imagined.[25] The point is that these philosophies, treated as scientific theories of mathematical practice, are naturalistically acceptable, while the above Simple and Glib philosophies are not.

I leave to others the task of determining which of these acceptable theories is best, but I should pause to take note of a disparity in the motivations behind various proposed philosophical theories of mathematics. Here, I have been arguing for a scientific, naturalistic approach that puts some restrictions on the range of acceptable philosophies and specifies that generally scientific criteria are applicable to the choice within that range. But recall the naturalist's claim that the role of extra-mathematical philosophy in mathematics is inspirational rather than justificatory. Some writers, usually mathematicians, philosophize about mathematics from this perspective, and it is not to be expected that an excellent inspiration will always be a scientifically adequate philosophy, or vice versa. A philosophy might, for example, be effective as an inspiration without even being adopted consistently; one philosophy might be better suited to one type of problem or one branch of mathematics, another to another. A philosophy might be effective as an inspiration even if it is 'justified' on unscientific grounds, for example, as direct communication from extra-terrestrial aliens. There is nothing wrong in the inspirational use of philosophy—indeed, there is much to be lauded—but the scientific naturalist is motivated by a different, standardly scientific goal.

The picture I have been painting begins with scientific naturalism, with science as the fundamental study. Mathematical naturalism extends scientific naturalism by treating mathematics as a separate undertaking, open to investigation by scientific methods, but not subject to methodological interference from that source. As this picture takes shape, it is natural to ask why mathematics merits this special treatment. What, for example, is to block an astrological naturalism, which holds that astrological methods are not subject to scientific criticism? A move in this direction might be welcome to the pluralist, but the scientific naturalist is likely to feel some discomfort. The trick, then, is to explain what singles out mathematics from the rest. From the scientific naturalist's point of view, that is, from the point of view of science itself, I think there are two conspicuous points of disanalogy between mathematics and, for example, astrology—points that justify disanalogous treatment.

First, the scientific naturalist notes that the domain of science includes all of spatio-temporal reality, the entire causal order, but that pure mathematics has nothing to say about this domain. Philosophical accounts of mathematics might say something about it—for example, that it does or does not include mathematical things—but mathematics itself, on the naturalist's model, does not. When questions of spatio-temporal location or causality do arise, what's involved is actually a mathematized physical object, for example, an impure set or a field. The question of how these mathematized descriptions of scientific reality work is a deep and important one, but the point remains that the domain of pure mathematics does not overlap the domain of science. Philosophical accounts of mathematics might impinge on the domain of science, but that sort of conflict is another matter, to be settled by rejecting the philosophy (if it is unnaturalistic) or by scientific methods (if the philosophy is naturalistic).

Astrology is another story. Webster's[26] defines it as:

> ... a pseudo-science claiming to foretell the future by studying the supposed influence of the relative positions of the moon, sun, and stars on human affairs.

In this sense,[27] astrology posits new causal powers and makes new predictions about spatiotemporal events, a clear incursion into the domain of science. As that incursion makes claims that are not supported by standard scientific methods, it is to be counted an invasion and deplored. The contrast with pure mathematics is stark.

This first disanalogy between mathematics and astrology may explain why the scientific naturalist will find it less troubling to follow the mathematical naturalist's admonition not to criticize mathematical methodology than to follow an analogous admonition from the astrological naturalist. However, it does not explain why the scientific naturalist should be particularly interested in giving an account of mathematics. Consider, then, the second disanalogy, from the scientific perspective, between mathematics and astrology: pure mathematics is staggeringly useful, seemingly indispensable, to scientific theorizing, and astrology is not. Thus, one part of understanding science, within science, is understanding what pure mathematics is, what it does for science when it is used in application, and why it does this job so well. This is a strong motivation for a naturalistic study of mathematics with no parallel in the case of astrology.

The final question in this line of thought is: why does the naturalist prefer the standpoint of science to begin with? Is not this just scientism? One fundamental goal of scientific naturalism is to provide a justification of scientific method as the best available way of achieving the goals of science.[28] But this, even if it were accomplished, would not establish that our science is the only reasonable science. As Quine puts the point:

> Might another culture, another species, take a radically different line of scientific development, guided by norms that differ sharply from ours but that are justified by their scientific findings as ours are by ours? And might these people predict as successfully and thrive as well as we? Yes, I think that we must admit this as a possibility in principle; that we must admit it even from the point of view of our own science, which is the only point of view I can offer. (Quine 1972, p. 181)

On the one hand, even our science tells us that our way of describing the world is probably not the only way, or even the only effective way; on the other hand, it is the only way we have, a fact which must ultimately carry the day.

Notice, finally, that even if we establish that our science is the best way we know of achieving our goals, even if we accept that we can do no better than to embrace the best available way of achieving our goals, we have not justified those goals themselves. A culture or a species with different goals would properly elect to proceed differently. But there is no way around this, and I think the scientific naturalist who shares the goals of science must and should be satisfied with this much justification and ask no more.

4 Conclusion

I have suggested that the spirit of naturalism is better served by a mathematical naturalism that treats mathematics on its own terms than by a science-only naturalism that subjects mathematics to external evaluation. Implementation of this mathematical naturalism requires some subtle techniques; I have tried to indicate how these can be developed and tested. Finally, though the mathematical naturalist sees properly philosophical questions as irrelevant to practical methodological decisions, there remains room for a naturalistic philosophical study of mathematics within science, a study that must be sensitive to the conclusions of the naturalistic methodologist of mathematics; I have tried to cast some light on these interconnections. My guiding hope is that mathematical naturalism better understood will be mathematical naturalism better loved.

Notes

[1] See (Maddy 1995), (Maddy 1996), (Maddy 1998*a*), and especially (Maddy 1997), where some themes of this paper are also developed in more detail.

[2] See (Maddy 1990, Chapter 4) for details.

[3] For the sake of completeness, it is worth noting a third, closely related difficulty, this one involving set-theoretic claims that do not present themselves as attractive axiom candidates, but which might, nevertheless, be regarded as welcome consequences of more appealing axiom candidates. (I have in mind, for example, $V \neq L$ or Projective Determinacy. See (Maddy 1998*b*).) The first problem is to explain what, if anything, legitimately disqualifies these from serious consideration as axiom candidates; the second problem is to explain what, if any, support they legitimately provide to axiom candidates that imply them.

[4] Examples of Simple Realism can be found in (Gödel 1964) and (Maddy 1990).

[5] For example, see Balaguer's 'full-blooded platonism' (1995).

[6] The contemporary development of this line of thought begins with Benacerraf (1973), though Benacerraf raises it as a question about Simple Realism, not as an argument against it. For discussion, see (Maddy 1990, §2.1), (Field 1989, pp. 25–30), and (Burgess 1990).

[7] Levy and Solovay (1967) show that CH is also consistent with and independent of ZFC plus various large cardinal axioms.

[8] The logic here is assumed to be first-order. The position sketched should not be confused with Hilbert's more complex and subtle view.

[9] Not to be confused with Balaguer's version, cited earlier. More subtle versions of these Glib views will be considered in §3, below.

[10] One version appears in Chihara (1973, Chapter 2), under the name 'Mythological Platonism'. This is not, however, the view that Chihara defends.

[11] This may sound like an idealization of actual practice, in the sense of the physicist's frictionless plane, but I think that this is the wrong analogy. What the physicist leaves behind are real causal factors; they are left out of the story to simplify it, so that deeper, more fundamental factors can be brought to the fore. But when a description of set-theoretic practice leaves out the philosophy (and other extra-mathematical considerations), the naturalist's position is that nothing truly functional has been left out; rather, distortions have been stripped away. A better analogy would be the removal of impurities to obtain a pure sample of a substance under study.

[12] I discuss this point in (Maddy 1996).

[13] We are focused here on the philosophical intruders, but there are others. For example, a mathematician's beliefs about the preferences of the NSF panel might influence her preference for one theory over another, but this influence, according to the naturalist, is not legitimately justificatory, and will also be set to one side.

[14] For example, in (Maddy 1996; 1998*a*; 1998*b*), and especially in (Maddy 1997).

[15] Actually, he argues that it would remain legitimate even if it were shown to be independent, and he predicts (correctly) that this will be shown.

[16] That is, Gödel's axiom of constructibility ($V = L$) and large cardinal axioms (LC).

[17] See (Maddy 1998*a*; 1998*b*), and especially (Maddy 1997).

[18] To say this, obviously, is not to deny that within any particular theory of sets, the sentence 'CH or not-CH' is true, that is, a theorem.

[19] See (Maddy 1996) for an example.

[20] This term is meant to signal and to set aside the complicated question, familiar to philosophers, of how far you have to look, how careful or how smart you have to be, for your neglect of a certain possibility not to undermine the claim that you are proceeding rationally.

[21] I have nothing to add to the debate on how these distinctions might be drawn.

[22] Here I am ignoring the possibility, noted earlier, that a principled distinction between sociology and mathematics could some day convict a mathematical community of erring by making their practice conform to some sociological (hence extra-mathematical) goal.

[23] A third distinction between the mathematical naturalist and the sociologist applies also to the scientific naturalist: leaving philosophical considerations aside, both naturalists undertake to separate legitimately justificatory arguments (like agreement with experiment or proof from accepted axioms) from other extra-scientific (extra-mathematical) pressures, like the preferences of grant-conferring agencies. Once again, I have nothing to contribute to the debate over this distinction.

[24] Though obviously not a science-only naturalist.

[25] Some who find Formalism or Fictionalism attractive as accounts of set theory may still be tempted to insist that the identities of simple arithmetic are robustly true, and I suspect that this position is defensible. Though I doubt that the role of mathematics in well-confirmed science can yield the general ontological consequences drawn in the well known indispensability arguments (see (Maddy 1995)), simple arithmetical identities do seem to be applied in uniquely literal form; perhaps these applications are even logical rather than fully mathematical. But admitting that $2 + 2 = 4$ is true in a sense that the Axiom of Choice is not, admitting even that the former is 'contentful' in a sense that the latter is not,

need not commit one to a Hilbert-type program of justifying the content-free in terms of the content-ful. To be motivated in that direction, one must also hold that pursuit of the content-ful is the overriding goal of mathematics, and this claim hardly squares with practice.

26. Webster's *New World Dictionary of the American Language*, second college edition, (Cleveland, OH: William Collins Publishers, 1979).

27. There are other interpretations of the goal and method of astrology. For example, discussions of an astrological chart between the astrologer and the subject are expressed in terms of archetypes that seem to go deep into human psychology; some astrologers hold that this process allows the subject to tap into a deeper level of understanding of his/her life and actions, which can be beneficial. On this interpretation, the astrologer is engaged in psychological counseling, not in describing causal mechanisms or making predictions. There remains room for debate within psychology about the effectiveness of this therapeutic technique, but this is a local discussion within science to be settled by standard scientific methods.

28. Note the similarity to the mathematical naturalist's assessment of mathematical methods.

Bibliography

Balaguer, M. (1995). A Platonist epistemology. *Synthèse*, **103**, 303–25.

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*, **70**, 661–80. Reprinted in *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 272–94. Cambridge University Press, 1983.

Burgess, J. (1990). Epistemology and nominalism. In *Physicalism in mathematics* (ed. A. Irvine), pp. 1–150. Kluwer Academic, Dordrecht.

Chihara, C. (1973). *Ontology and the vicious circle principle.* Cornell University Press, Ithaca, New York.

Field, H. (1989). *Realism, mathematics and modality.* Blackwells, Oxford.

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.

Hadamard, J., Baire, R., Borel, E., and Lebesgue, H. (1982). Five letters. Translations in and reprinted in G. Moore. *Zermelo's Axiom of Choice*, pp. 311–320. Springer-Verlag, New York. Originally published in 1905.

Levy, A. and Solovay R. M. (1967). Measurable cardinals and the continuum hypothesis. *Israel Journal of Mathematics*, **5**, 234–48.

Maddy, P. (1990). *Realism in mathematics.* Clarendon Press, Oxford.

Maddy, P. (1995). Naturalism and ontology. *Philosophia Mathematica*, **3**, 248–70.

Maddy, P. (1996). Set-theoretic naturalism. *Journal of Symbolic Logic*, **61**, 490–514.

Maddy, P. (1997). *Naturalism in mathematics.* Clarendon Press, Oxford.

Maddy, P. (1998*a*). $V = L$ and Maximize. In *Logic Colloquium '95* (ed. V. Harnik and J. A. Makowsky). (To appear.)

Maddy, P. (1998*b*). Progress in contemporary set theory. In *Mathematical Progress* (ed. H. Breger and E. Grosholz). (To appear.)

Putnam, H. (1972). Philosophy of logic. Reprinted in his *Mathematics, matter and method, Philosophical Papers,* Vol. 1 (2nd edn), pp. 323–57. Cambridge University Press (1979).

Quine, W. V. (1972). Responses. Reprinted in (Quine 1981, pp. 173–86),

Quine, W. V. (1975). Five milestones of empiricism. Reprinted in (Quine 1981, pp. 67–72).

Quine, W. V. (1981). *Theories and things.* Harvard University Press, Cambridge, Massachussetts.

Departments of Philosophy and Mathematics
University of California at Irvine
Irvine CA 92697
USA
email: pjmaddy@uci.edu

10

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 1998*c*).)

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 1998*c*).

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 theorists[1] 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 Constructibility ($V = L$) true?' or 'Do inaccessible cardinals exist?', and to believe that these are meaningful questions. (I am probably here describing the *Simple Realism* of (Maddy 1998*c*); 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 1998*c*), (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 logic[3]) 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 (1998*c*); 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 competing theories is not irrational; this is moving towards the *Subtle Formalism* of (Maddy 1998*c*).

The doctrine of *naturalism* is described in the previous chapter (Maddy 1998*c*); 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 XX$^{\text{th}}$ 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 XIX$^{\text{th}}$ 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–Fræenkel 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 XIX$^{\text{th}}$ 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) *$\cdot : G \times G \to G$ 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 XIX$^{\text{th}}$ 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 XIX$^{\text{th}}$ 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) $\langle \alpha x + \beta y, z\rangle = \alpha \langle x, z\rangle + \beta \langle y, z\rangle \quad (\alpha, \beta \in \mathbb{C},\ x, y, z \in H)$;

(ii) $\langle y, x\rangle = \overline{\langle x, y\rangle} \quad (x, y \in H)$;

(iii) $\langle x, x\rangle \geq 0 \quad (x \in H)$;

(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 $\theta : A \to B$ such that

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

is an algebra homomorphism and such that $\theta : (A, \|\cdot\|) \to (B, \|\cdot\|)$ 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

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

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

$$\int_0^1 |f(t)|^2 \, \mathrm{d}t$$

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

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

(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

$$\int_0^1 |f(t) - g(t)| \, \mathrm{d}t = 0\,;$$

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 a 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 Constructibility is true; see also (Maddy 1998*c*).

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 1998*c*)) 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 XIX$^{\text{th}}$ 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 XIX$^{\text{th}}$ 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

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

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 realization[14] 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. Moreover 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 $f : \mathbb{R} \to \mathbb{R}$. 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 + ¬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 + ¬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 XX$^{\text{th}}$ 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; this implies that $|\mathcal{P}(\Omega)| \leq |\Omega|$, contradicting 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 ZFC + CH and ZFC + ¬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 XX$^{\text{th}}$ 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 XX$^{\text{th}}$ 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. F. 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 $C(K)$. *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. (1998*c*). 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.

Moschovakis, Y. N. (1980). *Descriptive set theory.* North-Holland, Amsterdam.

Solomon, R. (1995). On finite simple groups and their classification. *Notices American Math. Soc.*, **42**, 231–9.

Solovay, R. M. (1970). A model of set theory in which every set of reals is Lebesgue measurable. *Annals of Mathematics*, **92**, 1–56.

Stewart, I. (1989). *Galois theory* (2nd edn). Chapman & Hall, London.

Woodin, W. H. (1998). The tower of Hanoi. *This volume*, 329–51.

Department of Pure Mathematics
University of Leeds
Leeds LS2 9JT
England
email: pmt6hgd@leeds.ac.uk

PART III

Realism in mathematics

11

A credo of sorts

V. F. R. Jones

Let me begin by claiming to be quite ordinary among mathematicians in my attitude to my subject, with little understanding of its philosophical underpinnings. I remember being worried by Russell's paradox as a youngster, and am still worried by it, but I hope to demonstrate, by a series of anecdotes and musings, that it is not at all difficult to live with that worry while having complete confidence in one's mathematics. Let us start on very very solid ground.

The Fourier transform of $f(x)$ is the function

$$\widehat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{\mathrm{i}kx} \, \mathrm{d}x.$$

Fourier inversion is the remarkable fact that

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \widehat{f}(k) \mathrm{e}^{-\mathrm{i}kx} \, \mathrm{d}k.$$

This property, known also as "unitarity", expresses the fact that nothing is lost about $f(x)$ by passing to $\widehat{f}(k)$. The Fourier transform is a great workhorse of mathematics, both pure and applied. Passing, as it does, between differentiation and multiplication, it can transform your problems out of existence in differential equations. It is at the heart of the uncertainty principle of quantum mechanics. It allows one to control subtle properties of the smoothness of functions. Fourier analysis splits complicated waves up into simpler components allowing us to understand both sound and light. The "Fast Fourier transform" can in some cases compute the outcome of a natural process faster than it is actually happening. Data can be adequately reconstructed from only a partial knowledge of its Fourier transform. The eigenfunctions of the Fourier transform are the Hermite functions, themselves of no small interest—the first is the Gaussian. All of harmonic analysis on groups begins with the Fourier transform. Wavelets provide alternatives to the Fourier transform which may work better for specific kinds of function, but surely the "uniformly best transform" is the Fourier transform.

To doubt the "truth" of the Fourier transform, however the word truth be interpreted, would be mathematical lunacy. It stands impregnable to any attacks of logical inconsistency or contradiction. I believe that Fourier himself gave several proofs of the convergence of a Fourier series, some correct, some wrong.

To emphasize my point even further, let me recall a story from my undergraduate days.

As a physics student I once took an introductory course on quantum optics. One of the first topics on the agenda was, naturally, the Fourier transform, and our enthusiastic mentor undertook a proof of Fourier inversion. His proof was totally wrong. Not just technically, but utterly, with no redeeming features. The professor, however, was blissfully unconcerned and, looking back on it, went on to show a more than useful intuition for the Fourier transform.

Of course, a solid proof is part of mathematics and the main significance for me of this event was to launch me on my journey away from physics and towards mathematics, where I felt I would never have to put up with such intellectual outrage. For mathematics is seductive to the young mind in that the decision to accept an assertion as correct is entirely up to the individual pondering the question. There is no outside authority who can decide the matter. No need to draw authority by quoting Kant, Frege, or some other old luminary. It is within oneself, independent of age, culture, stature, maturity, or experience, that the decision is to be made.

A less subjective significance of the Fourier inversion anecdote is in showing that the "truth" of a great piece of mathematics amounts to far more than its proof or its consistency, though mathematics stands out by requiring as a *sine qua non*, a proof that holds up to scrutiny. Our physicist, on the other hand (and I do not claim that all are like him), could not have cared less about how to prove Fourier inversion. He knew it was true from the tangible evidence of his science.

Here is a further story along similar lines. A certain physicist, who had made major contributions to conformal field theory (where analytic functions are the source of the rich structure), was giving an "introductory" lecture on the subject. One of his arguments involved a dubious use of analytic functions and a member of the audience (also a physicist in fact) made a series of queries ending with "So you're talking about a non-constant holomorphic function with compact support?" The answer: "Yes, of course." Here the physicist did not have the direct *experimental* verification of his conclusions as in the case of the Fourier transform, but was relying on the richness of the structure he was describing and the consensus of his fellow workers to free him from worry over this absurdity.

In both cases the physicist's reaction on being told that his proof was riddled with holes would be essentially to say, "Oh that's just something for the mathematicians to worry about." So it is for the mathematician with respect to the logician. If the day ever comes when the logicians find some inconsistency in arithmetic, our reaction will surely be, "Oh that's just a trick of the logicians; let them worry about it." And one can almost hear the inconsistency coming—perhaps there will be a proof of the existence of a contradiction, but that contradiction is too long to even contemplate, so we may quite happily behave as if it did not exist. (I believe that Woodin (1998) raises a similar possibility, though much more carefully considered, in his chapter.) The inconsistency will come with its own disclaimer. It must. The Fourier transform cannot fall.

But one may ask questions of the Fourier transform. For what functions $f(x)$ does it work? What is the nature of the convergence of the integral? One is immediately in the heartland of analysis—the Hilbert space $L^2(\mathbb{R}, dx)$, the Schwarz space and distributions (themselves originally created to justify the physicists' fanciful δ-functions, such justification being regarded as little more than a joke by legions of physicists). All these extremely carefully thought out concepts are built on the real numbers. So the next question: what is a real number? Here a sense of unease begins as we know full well that our answer—an element of a complete ordered field of which there is only one up to isomorphism—is a poor rendering of one's feeling for the real numbers. And of course the real numbers are a set, are they not? And here, only two simple questions away from the rock-solid Fourier transform, we are in trouble, for what *is* a set? My own notion of a set is very primitive, certainly not going beyond "naive set theory".

One hundred years ago the romantic mathematician would have wanted to, and sometimes did, try to construct an edifice of impeccable logic starting from scratch and ending up with the Fourier transform and its kin. We must thank Gödel for freeing us from the bonds of this romantic desire, for we now know that any such attempt is doomed to failure. We must at some point say that we *believe* in the soundness of our mathematics in a way not at all dissimilar to religious belief, though I would guess that a poll would show the vast majority of mathematicians to be disbelievers in the theistic dogma of any conventional religion. Perhaps a better analogy is with evolution. It is impossible to "prove" that evolution of species has occurred, but if one is even slightly acquainted with the fossil record and other empirical evidence, a pattern so powerful emerges that it is clearly folly to deny that species are descended from others with quite different characteristics.

The mathematician is as certain of his faith in mathematics as he is in the fact that a ball will drop if held above the ground and released—more sure than that the sun will rise the next morning. I would like to illustrate the grounds for this faith with a few selected mathematical events with which I am familiar. I cannot expect to fully convince the doubter with a few vignettes, though. This mathematical faith is earned at least as much as it is given—earned by many years of work.

I have mentioned Hilbert space. With that we are close to von Neumann algebras, which are the best-behaved infinite-dimensional generalization of ordinary matrix algebra. A von Neumann algebra is a family of continuous linear maps from a Hilbert space $\mathcal{H}$ to itself, which is closed under addition, multiplication, the adjoint $*$, and convergence, where $\langle a^*\xi, \eta\rangle = \langle \xi, a\eta\rangle$ defines a^*, and a_n converges to a if $a_n\xi \to a\xi$ in the norm $\|\cdot\|$, ξ and η being any vectors in $\mathcal{H}$. (The inner product in the Hilbert space is denoted by $\langle\cdot,\cdot\rangle$, and the norm is defined by

$$\|\xi\| = \sqrt{\langle \xi, \xi\rangle} \quad (\xi \in \mathcal{H}).)$$

We write $\mathcal{B}(\mathcal{H})$ for the family (actually, C^*-algebra—see (Dales 1997)) of all continuous linear maps from $\mathcal{H}$ to itself.

Why should we study von Neumann algebras? Given that we have decided our whole house is built on faith, this becomes a vital question. Not any mathematical structure should be as worthy of study as any other. It would not suffice to say that von Neumann algebras are a generalization of matrix algebras. In fact there are many motivations—ergodic theory, group representations and, for me by far the most important, the mathematical structure underlying quantum mechanics, where the states of a system are rays in some Hilbert space and observables are linear operators.

Although it requires knowledge of some elementary facts about Hilbert space, I would like to give the proof of von Neumann's celebrated "density" or "bicommutant" theorem as it was this beautiful result, as much as any motivation, that began study in the area. The theorem (von Neumann 1929) asserts that a von Neumann algebra M (containing the identity operator I) is equal to its own double commutant M'', where the commutant S' of a set S of linear operators on $\mathcal{H}$ is

$$\{a \in \mathcal{B}(\mathcal{H}) : as = sa \quad (s \in S)\}.$$

Since M is topologically closed, it suffices to show that M is dense in M'', which is obviously closed, any commutant being closed. So we let a be in M'', and choose a neighborhood $V = V(\varepsilon; \xi_1, \dots, \xi_k)$ of a, where ε is a real number > 0 and $\xi_1, \dots, \xi_k$ are vectors in $\mathcal{H}$. The set V is

$$\{b \in \mathcal{B}(\mathcal{H}) : \|a\xi_i - b\xi_i\| < \varepsilon \quad (i = 1, \dots, k)\}.$$

Let ξ be the vector $\xi_1 \oplus \xi_2 \oplus \dots \oplus \xi_k$ in $\mathcal{H} \oplus \mathcal{H} \oplus \cdots \oplus \mathcal{H} = \mathcal{K}$. The von Neumann algebra

$$\widetilde{M} = \{x \oplus x \oplus \cdots \oplus x : x \in M\}$$

acts on $\mathcal{K}$. We consider the *closure* in V of the subspace $\widetilde{M}(\xi_1 \oplus \xi_2 \oplus \cdots \oplus \xi_k)$. Since $\widetilde{M}$ is a $*$-algebra, the projection onto V commutes with $\widetilde{M}$ and, by a simple matrix computation, this projection is given by a matrix with entries in M'. Since $a \in M''$, $a \oplus a \oplus \dots \oplus a$ also commutes with this projection, so that

$$(a \oplus a \oplus \cdots \oplus a)\widetilde{M}(\xi_1 \oplus \xi_2 \oplus \cdots \oplus \xi_k)$$

is in the closure of $\widetilde{M}(\xi_1 \oplus \xi_2 \oplus \cdots \oplus \xi_k)$. Applying this to the identity element of $\widetilde{M}$, we deduce the existence, for our given ε, of an element b of M with

$$\|(b\xi_1 \oplus b\xi_2 \oplus \cdots \oplus b\xi_k) - (a\xi_1 \oplus a\xi_2 \oplus \cdots \oplus a\xi_k)\| < \varepsilon/k\,.$$

This ends the proof. I shall discuss this proof later in the talk.

The simplest von Neumann algebras are "factors", i.e., their centres (those elements which commute with all others in the algebra) consist only of scalar multiples of the identity. The most obvious factor is the algebra $\mathcal{B}(\mathcal{H})$ of all operators, elements of which are matrices (with respect to some orthonormal basis) with some growth condition on the size of the matrix entries. The *trace* is the partially defined function $\mathcal{B}(\mathcal{H}) \mapsto \mathbb{C}$ given by the sum of the diagonal elements, provided that sum converges appropriately. Projections are the operators which can be written as matrices

$$\begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & \ddots & & & & \\ & & & & 0 & & & \\ & & & & & 0 & & \\ & & & & & & 0 & \\ & & & & & & & \ddots \end{pmatrix}$$

with respect to some orthonormal basis. Clearly the trace of a projection is the dimension of its range space.

Murray and von Neumann (in the 1930s) made the remarkable discovery that there are factors $\mathcal{R}$ of an entirely different nature, the so-called type II_1 factors, which are infinite-dimensional but admit a trace tr which is a linear function, defined everywhere, for which

$$\operatorname{tr}(ab) = \operatorname{tr}(ba) \quad (a, b \in \mathcal{R})$$

and for which the trace of a projection, in contrast to the $\mathcal{B}(\mathcal{H})$ case, can be any positive real number (though one may normalize the trace so that the set of traces of projections is the unit interval [0,1]). This brought a notion of continuous dimension into mathematics, and remains one of the main seductive features about von Neumann algebras. The study of general factors has been greatly advanced by many people, most notably Tomita, Takesaki, and Connes, and in some sense all factors other than $\mathcal{B}(\mathcal{H})$ itself can be obtained relatively easily from a II_1 factor at their core. A lively interaction with mathematical physics has more than justified the original motivation for the study of von Neumann algebras from the mathematical foundations of quantum mechanics.

Given the interest of factors M one might want to study subfactors N of a given factor M. Ironically, although this has proved an extremely fruitful area, the *a priori* justification was rather weak—flying in the face of our insistence on serious motivation for the study of von Neumann algebras. Progress in mathematics will never follow any rules imposed upon it. Returning to our situation $N \subset M$ of II_1 factors, the projection operator from M to N becomes a projection in a II_1 factor, and one can think of its trace (or rather than inverse of its trace because of the normalization) as giving the "dimension" of M as a "vector space" over N. We call this number the *index* of N in M, written $[M : N]$. For instance if M were the $(k \times k)$-matrices over N, we would have $[M : N] = k^2$.

The Murray–von Neumann theory would suggest that any real number ≥ 1 can be the index $[M : N]$ for an appropriate $N \subset M$. In fact this is not the case.

The allowed values of $[M : N]$ less than 4 are precisely the numbers

$$4\cos^2 \pi/n$$

for $n = 3, 4, 5, 6, \ldots$ But from 4 onwards, continuous dimensionality comes into play and one may obtain any real number ≥ 4.

This is a striking result which needed careful proof. There is to date essentially only one proof of the result (there are several variations, all quite superficially equivalent). But from the beginning it seemed likely that the result would be true. The very surprising nature of the conclusion suggested that the arguments to prove it would be justified. And although there has been an undisputed proof from the beginning, I personally would still have trouble believing the result were it not for circumstantial evidence, such as the wealth of structure and interrelations that have grown up around it. This is the point in this chapter where I perhaps come closest to saying something original.

Proofs are indispensable, but I would say they are necessary but not sufficient for mathematical truth, at least truth as perceived by the individual.

To justify this attitude let me invoke two experiences of current mathematics, which very few mathematicians today have escaped.

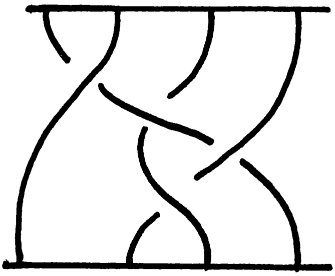
The first is computer programming. To write a short program, say 100 lines of C code, is a relatively painless experience. The debugging will take longer than the writing, but it will not entail suicidal thoughts. However, should an inexperienced programmer undertake to write a slightly longer program, say 1000 lines, distressing results will follow. The debugging process becomes an emotional nightmare in which one will doubt one's own sanity. One will certainly insult the compiler in words that are inappropriate for this essay. The mathematician, having gone through this torture, cannot but ask: "Have I ever subjected the proofs of any of my theorems to such close scrutiny?" In my case at least the answer is surely "no". So while I do not doubt that my proofs are correct (at least the significant ones), my belief in the results needs bolstering. Compare this with the debugging process. At the end of debugging we are happy with our program because of the consistency of the output it gives, *not* because we feel we have proved it correct—after all we did that at least twenty times while debugging and we were wrong *every time.* Why not a twenty-first? In fact we are acutely aware that our poor program has only been tested with a limited set of inputs and we fully expect more bugs to manifest themselves when inputs are used which we have not yet considered. If the program is sufficiently important, it will be further debugged in the course of time until it becomes secure with respect to all inputs. (With much larger programs this will never happen.) So it is with our theorems. Although we may have proofs galore and a rich surrounding structure, if the result is at all difficult it is only the test of time that will cause acceptance of the "truth" of the result.

The second experience concerning the need for supplements to proof is one which I used to dislike intensely, but have come to appreciate and even search for. It is the situation where one has two watertight, well-designed arguments that lead inexorably to opposite conclusions. Remember that research in mathematics involves a foray into the unknown. We may not know which of the two conclusions is correct or even have any feeling or guess. Proof at this point is our only arbiter. And it seems to have let us down. I have known myself to be in this situation for months on end. It induces obsessive and anti-social behaviour. Perhaps we have found an inconsistency in mathematics. But no, eventually some crack is seen in one of the arguments and it begins to look more and more shaky. Eventually we kick ourselves for being so utterly stupid and life goes on. But it was no tool of logic that saved us. The search for a chink in the armour often involved many tricks including elaborate thought experiments and perhaps computer calculations. Much structural understanding is created, which is why I now so value this process. One's feeling of having obtained truth at the end is approaching the absolute. Though I should add that I have been forced to reverse the conclusion on occasions...

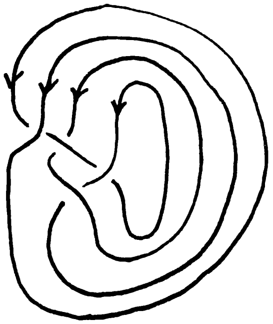
Let us return to the result that led us here, namely the surprising conclusion about the index of subfactors. It is not a particularly elementary result, so I will leave it behind to explore one of the outcomes of it, which is one of the structural items enhancing the truth of the subfactor result. It is indeed an elementary result and because of the quirks of mathematical history, at least as striking. I refer to the discovery of a new polynomial invariant for knots (see (Jones 1990)).

By *knot* I mean a smooth closed curve in $\mathbb{R}^3$, and two knots are considered the same if they can be smoothly deformed from one to another. A *link* is a collection of disjoint knots. The smooth deformation corresponds precisely to the intuitive notion of moving a real piece of string around in space. This has the interesting effect of bringing simple physical intuition to bear on the mathematics, something quite absent in von Neumann algebras. I would like to illustrate by giving a proof of a simple result due to Alexander in the 1920s (Alexander 1923). It was explained to me by Joan Birman within five minutes. A *braid* is a system of curves in $\mathbb{R}^3$ which begin on a certain horizontal plane and end at points on another plane, directly below the starting point, the key defining property being that the curves never have a horizontal tangent vector.

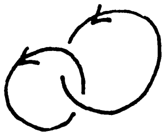
Here is a picture of a braid where, as is traditional, all the points at the top lie on a straight line:



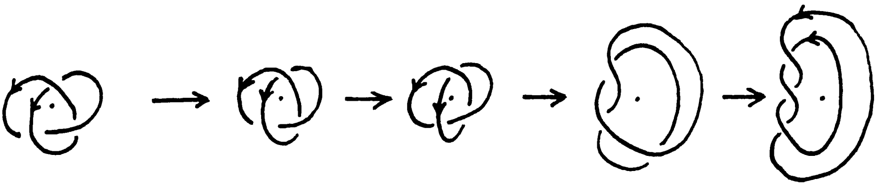
We have given an informal definition, but the notion of a braid is readily formalized, the most difficult part being to say exactly when two braids are "the same". As with knots, that corresponds to an intuitive physical notion of smooth motions between the two horizontal planes. One may take a braid and form a knot (or link) by tying the tops to the bottoms. For the above braid the link which is its "closure" is depicted below.



The link obtained in this case is visibly rather simple and could be redrawn as follows.

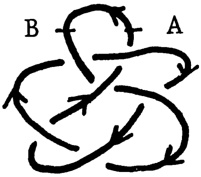


Note that closed braids come with an orientation inherited from that of the braid, as shown. The result of Alexander is that any oriented knot or link can be obtained as the closure of some (highly non-unique!) braid. To prove this result requires two observations. The first is that if there is a point somewhere on the diagram of a knot, around which the orientation of the knot moves in a consistently clockwise (or anticlockwise) direction, the job is done. Just open the knot on any ray coming out from the central point and stretch that part so that it becomes the straight "back" of the closed braid, as in the next diagram.

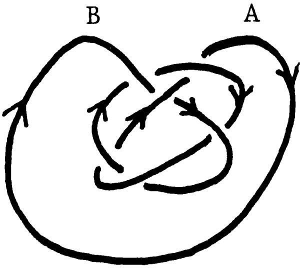


The central point (or rather a straight line through it perpendicular to the paper) is called the "braid axis".

To complete the proof one simply has to show that there is always a braid axis. Unfortunately this is not always the case, as we see in the next picture.



What we can do is choose *any* point in the plane and rearrange the picture so that it becomes a braid axis. This is done by going around the knot until one finds a stretch that is going the wrong way. One then isolates a short stretch going the wrong way and "throws it over one's shoulder" until it is on the other side of the knot, going around correctly. For the above example the only bad stretch is between A and B and after throwing it over one's shoulder one gets the picture below



The picture above is now a closed braid. The only thing that could go wrong in the above procedure is that, in trying to throw a bit of string over one's shoulders, one may meet a crossing. This is easily handled. Since we are proceeding one short stretch at a time around the knot, simply isolate that crossing and, if it happens to prevent our throwing over our shoulders, throw it the other way. When we have arrived back at the beginning point, we see a closed braid. This ends the proof.

Historically, this "knot-to-braid" process played a major role in the discovery of the knot polynomial in (Jones 1985), but that is not the point I want to make here, though one could say a lot about how it reinforced both one's belief in the index result and some results about knots and braids. Rather I would like to compare and contrast the two proofs—of von Neumann's density theorem and Alexander's closed braid theorem, call them Theorem vN and Theorem A respectively.

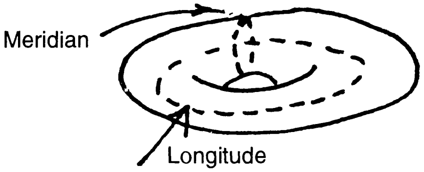
Theorem vN is a difficult theorem. To even understand what it is saying requires considerable mathematical background. To properly understand the proof takes many hours, and to have discovered the result and its proof was a major achievement. Theorem A, on the other hand, is easy. The result and its proof could be explained quite rapidly to a clever high school student. Yet a careful analysis of these proofs reveals that the proof of Theorem A, if properly formalized, would be much longer than that of Theorem vN. One would have to be precise about the kinds of continuous deformations that are allowed, and constructing the functions required for the "throwing over the shoulder" trick would be a nightmare. So why do we consider Theorem A to be so easy? The answer is that Theorem A concerns a very concrete situation, and we are able to bring to bear our full intuition about three-dimensional space on the problem. If we were two-dimensional creatures then proving this theorem would be another story entirely and would require much more *formal* argument. Theorem vN is, by contrast, infinite-dimensional and we cannot rely on simple three-dimensional intuition, so a formal proof is required. All the details of the proof, going back to the definition of Hilbert space, would require only a couple of pages.

One of the interesting consequences of the use of three-dimensional intuition is that the field of low-dimensional topology has advanced in a way that is significantly different from other branches of mathematics. One is expected to "see" results in this field, and once the result, or partial result, has been "seen", it requires no further discussion. I do not wish to criticize this approach. I have myself "seen" several results in this field, and believe them to be as correct as any other mathematics.

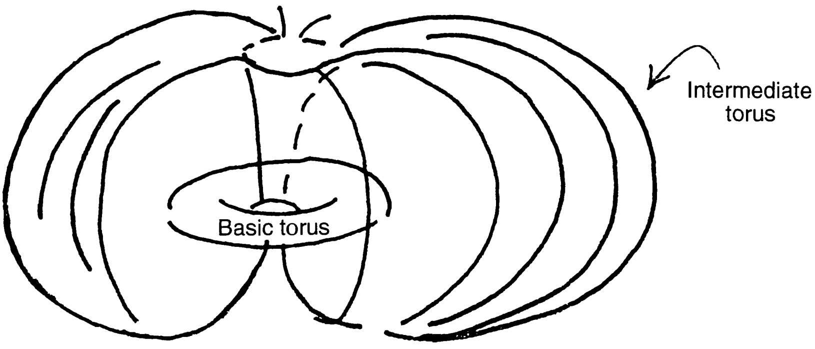
Here is an example which is the first significant "seeing" requirement—the point of entry into low-dimensional topology. The three sphere S^3 is

$$\{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\}.$$

It cannot be embedded in $\mathbb{R}^3$ so is not wholly concretely visualizable, but almost—if one point is removed from S^3 it becomes the same as $\mathbb{R}^3$ itself (this is readily seen by stereographic projection). Thus another way to approach S^3 is as the one-point compactification of $\mathbb{R}^3$. The result to which I refer is that S^3 is the union of two solid tori, joined along the boundary so that a meridian of one becomes a longitude of the other and vice versa.



To "see" this, think of a solid torus as a central circle surrounded by tori of increasing size up to the boundary torus. Now place a solid torus in $\mathbb{R}^3$ so that it lies on the (x, y)-plane with the z-axis through the middle. We can choose to add our point at infinity so as to compactify the z-axis to a circle. Now S^3 minus this circle is the same as $\mathbb{R}^3$ minus the z-axis and one can readily see a family of tori filling up $\mathbb{R}^3$ minus the z-axis, which give a solid torus when the compactified z-axis is added. Perhaps the following picture of one of these intermediate tori helps:



This is a very murky story from the formal mathematical point of view and the uninitiated will no doubt find it difficult, and my explanation inadequate. Fortunately there is a very simple formal description of the same picture: clearly $S^3 = T_1 \cup T_2$ with

$$T_1 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1,\ x_1^2 + x_2^2 \leq 1/2\},$$

$$T_2 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1,\ x_3^2 + x_4^2 \leq 1/2\}.$$

Then T_1 is obviously a solid torus since for each r with $1/2 \leq r < 1$, setting $x_1^2 + x_2^2 = 1 - r$ gives a circle and $x_3^2 + x_4^2 = r$ gives another so we get a family of tori, with the central circle being $r = 0$.

This formal picture (rendered even simpler using complex coordinates $z_1 = x_1 + \mathrm{i}x_2$, $z_2 = x_3 + \mathrm{i}x_4$) is complete but inadequate. If one does not "see" the other picture and how the tori, and the two circles, fill up the 3-sphere, one is not ready to take the next steps in low-dimensional topology. Of course this is just the beginning. There are more complex things to "see" and sequences of such visions are compounded one upon another in the same way as the elementary logical steps in a formal argument. If one "sees" the pictures, then one understands, but otherwise one cannot follow. In principle one could formalize the whole argument, but that would add nothing.

We see that mathematical "truth" in this field is very much contingent on our physical intuition and experience in three-dimensional space.

The reader not acquainted with low-dimensional topology may feel I am exaggerating for effect. Not at all. Once, at a seminar, one of the world's best low-dimensional topologists was presenting a major result. At a certain point another distinguished topologist in the audience intervened to say he did not understand how the speaker did a certain thing. The speaker gave an anguished look and gazed at the ceiling for at least a minute. The member of the audience then affirmed "Oh yes, I hadn't thought of that!" Visibly relieved, the speaker went on with his talk, glad to have communicated this point to the audience.

Such is truth in mathematics.

Bibliography

Alexander, J. (1923). A lemma on systems of knotted curves. *Proc. National Academy Sciences*, **9**, 93–5.

Dales, H. G. (1998). The mathematician as a formalist. *This volume*, 181–200.

Jones, V. F. R. (1985). A polynomial invariant for knots via von Neumann algebras. *Bull. American Math. Soc.*, **12**, 103–11.

Jones, V. F. R. (1990). Knot theory and statistical mechanics. *Scientific American*, **262**, 98–103.

von Neumann, J. (1929). Zur Algebra der Funktionaloperatoren. *Mathematische Annalen*, **102**, 340–427.

Woodin, W. H. (1998). The tower of Hanoi. *This volume*, 329–51.

Department of Mathematics
University of California
Berkeley
CA 94720–3840
USA
email: jones@math.berkeley.edu

12

Mathematical evidence

Donald A. Martin

1 Introduction

I will be discussing the subject of this conference, truth in mathematics. I will not, however, talk about the nature of mathematical truth but only about what counts as evidence for mathematical truth. I will do this not in general but only by way of two nearly thirty-year old examples from descriptive set theory, examples that seem similar to typical cases of evidence in empirical sciences. (The examples are not, of course, intended to be typical examples of mathematical evidence.) I hope it will be clear that my examples are cases of strong scientific evidence for the truth of mathematical propositions. I will raise and discuss the issue of whether there are stronger canons of evidence in mathematics, canons that my examples do not satisfy.

From the start, I want to make clear that I have nothing to say about the ontology of mathematics: about whether the subject matter of mathematics is platonic objects or whether it is other things, such as structures, concepts, or proofs. I also have nothing to say about the semantics of mathematical discourse. Nor do I have anything to say about whether evidence for accepting mathematical assertions is evidence for taking them as realistically true or is merely evidence for adopting them in some conventional or methodological sense. In short, I have nothing to say about the content of mathematical propositions.

This may appear foolish. How can one hope to decide what is evidence for a mathematical proposition until one knows what the proposition is about—what the content of the proposition is? For surely the answer to the content question can have great significance for questions about evidence. If, for example, mathematics concerns abstract objects with which we have no causal interaction, then one might—as did Paul Benacerraf in his famous paper (1973)—doubt that there can be any source of evidence for the truth or falsity of mathematical propositions.

The first and most important reason for my not discussing ontology and semantics is that I have no good ideas or even opinions about these questions. I confess that they mystify me. On the one hand, it is hard to avoid—in any account of mathematics that is at all realistic—an appeal to some kind of math-

Research partially supported by NSF Grant DMS-9505153. The author is indebted in very different ways to Joseph Almog, Penelope Maddy, and Yiannis Moschovakis.

ematical objects. On the other hand, there is a strong intuition that objects do not really matter. I remain impaled on the horns of this dilemma.

Another, and perhaps more acceptable, reason for studying epistemology while ignoring ontology and content is that we seem to have more intuitions, and—in the intersubjective sense, at least—more reliable intuitions, about mathematical truths and evidence for them than we do about the other questions. I hope the two examples presented later will illustrate this. Another illustration from the field of set theory is provided by the fact that set theorists with widely divergent philosophical views (formalists and platonists, say) often agree about what set-theoretic axioms are plausible or are reasonable to adopt, even when they are in complete disagreement about the significance of the adoption.

What I have just said about mathematics applies also to empirical sciences like physics. There is a disagreement among physicists—and even more among philosophers of science—as to whether physical theories are to be interpreted realistically. Sometimes such disagreement influences views about directions of research and about which theories should be accepted, but not all that often.

My hope is then that we can simply follow our intuitions about truth and evidence, bracketing all questions about ontology and content. When, in the sequel, I talk about how we obtain evidence for truth, platonists may—if they must—take me to be talking about how we discover facts about abstract mathematical objects. Formalists may take me to be talking ultimately about the proper grounds for choosing our axioms and so for deciding what is to count as true. But the frame of mind I prefer is one that puts aside issues about platonism, formalism, etc., considering truth and evidence in a direct and unanalyzed way.

2 Proof and axioms

What then does count as mathematical evidence? There is, of course, an obvious answer to the question of how one can come to know the truth of a mathematical proposition: namely, proof. Indeed, this may seem the *only* way to establish mathematical truth. When one says that a mathematical proposition is known, does not one just mean that it has been proved? And when one says that a mathematical question is open, does not one mean that no one has proved that the question has a positive answer or that it has a negative answer?

What is it to give a mathematical proof, a rigorous mathematical proof, of some statement S? It is rather surprising that such a clear and unambiguous standard answer is available to this question. To prove S, one must show that S follows by pure logic from the basic principles of mathematics. It is one of the triumphs of modern logic that one can say precisely what 'pure logic' is, in the relevant sense: namely, first-order logic. And it is a rather surprising fact that one can say precisely what the 'basic principles of mathematics' are: namely, the Zermelo–Frænkel (ZFC) axioms for set theory.

I do not want to make too much of the fact that the basic principles are axioms of set theory rather than of some other subject. It is simply a fact about current mathematics that all mathematical concepts can be defined in terms of sets (a minor qualification needs to be made for large objects such as categories) and that something is counted as a mathematical theorem if and only if it has been shown to follow logically from the ZFC axioms. I do not mean to imply that the ZFC axioms are *the* axioms of mathematics. Indeed, I think that further axioms are needed. I do not even mean to imply that a mathematician need know the ZFC axioms, at least not all of them. I do not mean to imply that set theory provides *the* foundation of mathematics. The alternative foundational notions that have actually been suggested do not provide anything essentially different, but perhaps a really different, richer foundation can be found. I do not want to suggest that I think that the division of mathematical knowledge into axioms and proved statements has ultimate significance. Finally, I do not mean to belittle the fact that some mathematicians think that the notion of mathematical proof should, in some way or other, be more restrictive than the standard notion.

The standard notion of mathematical proof that I have described provides us directly with examples of mathematical assertions that are accepted without proof: the ZFC axioms. (I ignore the degenerate sense of *proof* according to which each axiom can be proved by deducing it from itself.) Whatever is the evidence that leads to or led to the acceptance of the axioms, it is certainly not proof. Thus I will concentrate on the question of how one can come to accept—and to justify accepting—axioms of set theory.

One can raise this question about the current standard axioms, the ZFC axioms. How do we know they are true? Are they self-evident? In some cases, this seems implausible, at least on the surface. There was a bitter dispute about the Axiom of Choice, chronicled in (Moore 1982). Is Choice self-evident? What about Power Set, Replacement, and Infinity? If the axioms are not all self-evident, are there convincing arguments for their truth? If not, are they true by convention? I will make no general attempt to deal with this topic here. For the most part, I will be content to record, without justifying, my skepticism that the truth of each of the ZFC axioms is known with certainty. If my skepticism is misplaced, then some of what I say later will have diminished significance; so it matters that I am correct. For support I refer the reader to the discussion of the history of and evidence for the ZFC axioms in (Maddy 1988*a*). Maddy's account is intended 'to counteract the impression that these axioms enjoy a preferred epistemological status not shared by new axiom candidates'. (Considerations of the kind presented in (Woodin 1998) suggest that we lack certainty not just in set theory but even in number theory.)

There is, however, one important attempt to justify the ZFC axioms on which I feel impelled to comment, if only briefly. This attempt argues that the ZFC axioms are true of the intended concept of set, the *iterative concept of set*. (Gödel (1947) was perhaps the first to assert that the iterative concept implies the ZFC axioms, although see also (Zermelo 1930).) According to the iterative concept, sets are to be regarded as formed in a transfinite sequence of stages. At each stage is formed every possible set whose members were formed at earlier stages. The number of stages is supposed to be 'absolutely infinite'. Though this intuitive concept is obscure in several ways, it seems sharp enough to let one see that most or all the ZFC axioms are true of it. Consider, for example, the Axiom of Power Set. If a set x is formed at a stage s, then all members of s were formed at earlier stages. But then each possible subset of x—so each subset that is ever formed—is formed at stage s or earlier. Consequently the power set of x is formed at the next stage after x, and so x has a power set. Such arguments for justifying the axioms are important. I myself often present them when I teach courses in set theory. Nevertheless, I do not think that they establish with certainty the truth of the ZFC axioms. There are two ways to raise doubts about the arguments. One way is to admit that the iterative concept implies the axioms but to question whether that concept is a coherent one. One might, for example, question the consistency of ZFC. Another way to raise doubts is to argue that the iterative concept as ordinarily presented is really a combination of (1) a concept and (2) alleged facts about that concept. It would be possible to give pure formulation of the underlying iterative concept (whose core idea is simply that sets are formed from sets that have already been formed), a formulation from which Power Set, Infinity, etc., could not be deduced. (The first person I heard argue in this second way was Byeong-Uk Yi.)

Even the question of why the axioms *are* accepted, as opposed to that of why they *should be*, is a difficult one. The very fact—a sociological fact—that they *are* the standard axioms, and have been for quite some time, increases the difficulty of evaluating their epistemic status. For individual mathematicians, acceptance of an axiom is probably often the result of nothing more than knowing that it is a standard axiom. It would be easier to investigate the evidence for the ZFC axioms if we could move back in time to the period when they had been put forward, but had not yet become official dogma. Of course, we can read what people of that epoch said on the matter. Instead of doing this, though, I will take another route.

A standard complaint about set theory is that its concepts and methods are not sufficiently related to those of central parts of mathematics. This complaint is raised against set theory as a foundation for mathematics, and it is also given as a criticism of set theory as a branch of mathematics in its own right. In both cases I have doubts about the justice of the criticism. For my present purposes, however, there is an important way in which the complaint is true, and this—oddly enough—is a way in which set theory's location on the fringe of mathematics makes it more relevant to my concerns. The ZFC axioms are a largely adequate basis for current mainstream mathematics. But they are far from an adequate foundation for set theory itself as a mathematical discipline. The famous independence results of Gödel and Cohen showed that the ZFC axioms do not yield an answer to one of the most basic and simple questions of set theory: that of Cantor's Continuum Hypothesis. Moreover the methods they introduced, especially Cohen's method of forcing, have been used to show that a vast number and vast variety of set-theoretic sentences are consistent with and independent of the ZFC axioms. By now it is almost a major event in set theory if something is shown to follow from these axioms.

Now it is true that the independence phenomenon is not limited to set theory itself. In most parts of mathematics, analysis, algebra, even number theory, significant propositions have been proved consistent with and independent of ZFC. Mathematicians in these disciplines can usually console themselves that the propositions in question have a set-theoretic or a metamathematical character, and so lie on the fringe of their discipline. This is not always the case. Nevertheless, most mathematicians can, at least so far, not worry very much about the incompleteness of ZFC.

Within set theory, the matter is quite different. As a set theorist, one can spend one's time proving independence results. Admittedly, many set theorists are content to do this, and some of them consider proving independence results the main business of set theory. But if one wants outright answers to the important questions of set theory, then one must find new axioms that go beyond those of ZFC. Indeed, there has been considerable interest in new axioms for quite a long time, and some candidates have come forth. One prominent class of such candidates is the class of *determinacy hypotheses.* A substantial body of alleged evidence for determinacy hypotheses has been uncovered. Looking at this alleged evidence may be a way to shed some light on the question of what should and what should not count as good evidence for fundamental axioms. And there are advantages in considering axiom candidates rather than the standard ZFC axioms. The very fact that the former are not standard axioms allows one to evaluate their status in a cleaner way.

It should be admitted immediately that determinacy hypotheses (or the related *large cardinal axioms*, of which I will speak briefly later) are not likely in the near future to be adopted as full members of the canonical list of axioms—at least, not by the general mathematical community—however strong the evidence in their favor might be. As long as the ZFC axioms are more or less adequate for mainstream mathematics, the mathematical community will feel no need for new axioms. (Recall that perhaps the most influential of Zermelo's arguments for the Axiom of Choice was that the axiom was indispensable.) This means that the evidential requirements for adoption of new axioms are *de facto* very strong indeed, and the evidence for determinacy hypotheses does not fulfil these requirements. This may lessen the significance for current mathematics of the discussion that follows, but I do not see that it destroys the relevance of that discussion to the basic epistemological questions.

3 Mathematically proper evidence

In discussing mathematical evidence, it *may* be important to distinguish evidence simpliciter from what I will call *mathematically proper* evidence. By mathematically proper evidence for a mathematical proposition I mean evidence that counts toward giving *mathematical* knowledge of the truth of the proposition. Here 'mathematical knowledge' is intended to mean something like 'knowledge obtained in the mathematical way' or 'knowledge that makes the proposition count as mathematically "known".'

Remark. The term 'proper' is chosen for lack of a better one. I certainly do not want to imply that there is any lack of propriety in evidence to which the term does not apply.

Examples of mathematically proper evidence might be (a) proof, and (b) direct intuitions of truth, for example, intuitions of the truth of the principle of mathematical induction and of some of the axioms of set theory. I will later mention other, more problematic, examples.

For a possible example of evidence that is not mathematically proper, consider the contention of some mathematicians that the four-color theorem has not been mathematically established, because a physical system—a computer—has been relied on for part of the proof. The alleged problem in this case is apparently that the knowledge provided by the computer-assisted proof is not fully *a priori*. Surely many mathematicians do think that proper mathematical knowledge must be *a priori*. Someone may allege that computer-assisted proofs suffer from a second defect: a failure to provide certainty. Computer proofs, though, are no more susceptible to error than human ones; so it is hard to see the force of the allegation. Perhaps it is just a reformulation of the complaint about lack of *aprioricity*.

Some other kinds of evidence that may not qualify as mathematically proper are (a) inductive evidence for generalizations obtained by verifying instances, and (b) proofs establishing that statements hold with high probabilities. To avoid mixing up two different issues, let us consider (a) and (b) only in cases where the evidence itself is known in a mathematically proper manner (in particular, where there is no use of computers). Then what the evidence fails to convey seems to be *certainty*. Certain knowledge of the evidence does not yield certain knowledge of the proposition for which it is evidence. Is giving certainty then necessary for evidence to be mathematically proper? This cannot be true, at least if I was correct earlier in saying that some of the axioms of set theory are not known with certainty. Nor does something like 'high degree of certainty' help, since this will not eliminate the probabilistic proofs.

The discussion so far has turned up two candidates for necessary conditions that a kind of evidence be mathematically proper: being known *a priori* and yielding certainty. The former seems a plausible necessary condition, but the latter does not. In the last section of this chapter, I will examine a third candidate for a necessary condition, a candidate I will also find wanting. Though I am sure that the notion of mathematically proper evidence has at least a sociological significance, I will in the end be dubious about the real significance of the notion, and I am even dubious about whether there is a satisfactory way to characterize the notion.

The two examples of mathematical evidence that I will later describe do not provide certainty, but they are fully *a priori*. Thus they are more like the inductive and probabilistic cases than like the computer-proof case. They differ from these kinds of examples in two important ways.

First, they provide evidence, not just for isolated mathematical statements but for candidates for fundamental axioms. Thus they raise the general question: What kind of evidence is and should be relevant, or even decisive, for a proposition's being adopted as a fundamental axiom of mathematics? This is a question about what counts as a particular sort of mathematically proper evidence.

The second way my examples differ from the ones mentioned above is that with them much more is involved than, say, mere inductive confirmation. They provide a richer kind of evidence that is analogous to evidence for general theoretical statements in empirical sciences. I suspect that there are examples with this property in other branches of mathematics, but such examples are surely ones of evidence for propositions that one expects the ZFC axioms to decide.

4 Determinacy

Since both my examples are about determinacy hypotheses, it is time to say what determinacy hypotheses are.

Determinacy hypotheses concern strategies for infinitely long games of perfect information. Games of this sort were first studied by Polish mathematicians in the 1930s. (See pages 113–117 of (Mauldin 1981).) They were rediscovered by Gale and Stewart (1953) in the 1950s. In such games, two players, I and II, take turns making moves and continue to do so forever:

$$\begin{array}{lccccccc} \mathrm{I} & a_0 & & a_2 & & a_4 \ldots & & \\ \mathrm{II} & & a_1 & & a_3 & & a_5 & \ldots \end{array}$$

By some criterion, it is 'then' decided which player has won. The notions of *strategy* and *winning strategy* for I or II are defined in the obvious way. A game is *determined* if one of the players has a winning strategy.

Zermelo proved that all *finite* games of perfect information are determined. Is this true of all infinite games also? It turns out that it is not.

Steinhaus, who had suggested the idea that all infinite games might be determined, and Mycielski, who had refuted this suggestion, found and presented (Mycielski and Steinhaus 1962) an interesting way to weaken the refuted hypothesis. The weakening is now called the *Axiom of Determinacy* (AD). The axiom AD says that all infinite games in which the players are required to choose a natural number at each move are determined.

The problem with AD is that it contradicts the Axiom of Choice. One can try to make a virtue of this, as did Mycielski and Steinhaus, by arguing that it contradicts only the bad consequences of Choice. Choice implies that there are pathological sets of real numbers, non-measurable sets, uncountable sets without perfect subsets, etc. As Mycielski and others showed, AD implies that sets with these pathological properties do not exist. Nevertheless, few mathematicians were anxious to give up Choice, which had become a well-entrenched axiom.

Soon people hit upon the idea that, while AD is false in general, it might be true of all simple or definable sets. (Indeed, something like this was proposed in the original (Mycielski and Steinhaus 1962).) To see what this means, consider a general game of the type to which AD applies: a game in which the players take turns playing natural numbers. How are the winning conditions for player I given? They are given by a set A of infinite sequences of natural numbers. If the play $a_0, a_1, a_2, \ldots$ belongs to A, then I wins; if it does not belong to A, then II wins. In fact—at least if we restrict ourselves to games in which any natural number may legally be played in any position—a set A completely characterizes a game. We can, and do, call this game G_A. The Axiom of Determinacy says that G_A is determined for every A. Restrictions of it say that G_A is determined for restricted classes of A. For classes of sets A characterized by some notion of definability or simplicity, no contradiction with Choice, i.e., no contradiction with ZFC, has been found or is expected.

Here are some examples of such classes of sets A.

Open sets. The set A is *open* if, whenever an infinite sequence x belongs to A, there is a finite initial part s of x such that every infinite extension of s belongs to A. This is easier to understand in terms of the game G_A. If A is open and player I wins a play of G_A, then during the play some finite position s arises at which I has already won the play. If I wins, then I wins in finite time.

Borel sets. The Borel sets form the smallest class of sets containing the open sets and closed under countable unions, countable intersections, and complements. The Borel sets are classified into a transfinite hierarchy: a set is at level α if it takes α steps to build it from the open sets (where we do not count taking complements as steps).

Projective Sets. There is an obvious way to define *open* and *Borel* for sets B of n-tuples of infinite sequences of natural numbers. A set B is *projective* if it can be obtained from a Borel set by finitely many applications of the operations of complement and projection. Here we say that a set B of n-tuples is the *projection* of a set C of $(n+1)$-tuples if B is defined by

$$(x_1, \ldots, x_n) \in B \leftrightarrow (\exists x_{n+1})(x_1, \ldots, x_n, x_{n+1}) \in C\,.$$

The projective sets are classified into a hierarchy by counting the number of projections needed in generating them from Borel sets.

Sets Constructible from $\mathbb{R}$. These are the sets belonging to the smallest inner model of ZF (not including Choice) that contains all the ordinal numbers and all the real numbers. The inner model in question is called $L(\mathbb{R})$. The sets constructible from $\mathbb{R}$ form a much wider class than the class of projective sets.

Let *Open Determinacy* be the assertion that G_A is determined for every open A. Similarly, we define *Borel Determinacy* and *Projective Determinacy.* The assertion that G_A is determined for every set constructible from $\mathbb{R}$ is called $\mathrm{AD}^{L(\mathbb{R})}$.

Open Determinacy and even Borel Determinacy can be proved in ZFC. (See (Gale and Stewart 1953) and (Martin 1975).) But Projective Determinacy, even for the first level of the projective hierarchy, is not provable in ZFC. (This follows from a result of (Davis 1964).)

Starting in the 1960s, researchers discovered that determinacy hypotheses—in particular, Projective Determinacy and $\mathrm{AD}^{L(\mathbb{R})}$—have many important consequences. The hypothesis Projective Determinacy was used to settle almost every important question about projective sets. Developing the consequences of determinacy hypotheses became a very active field and remained so throughout the 1970s. Much of this work is described in (Moschovakis 1980) and in (Kechris 1995). By the end of the 1980s, another kind of support for Projective Determinacy and $\mathrm{AD}^{L(\mathbb{R})}$ was found: they were deduced in (Martin and Steel 1989) and (Woodin 1988) from so-called *large cardinal axioms* and were shown (mainly by Woodin) to imply large cardinal axioms, in a sense. Large cardinal axioms form another class of candidates for new axioms of set theory. The results just mentioned show that, over a wide range, large cardinal axioms and determinacy hypothesis are *one* class of axiom candidates, not *two* classes.

Remark. When I speak of 'determinacy hypotheses', I always mean assertions like Projective Determinacy and $\mathrm{AD}^{L(\mathbb{R})}$, assertions that do not contradict the Axiom of Choice. I never mean AD itself, which was never taken seriously as an axiom candidate.

In the next two sections, we will look at two particular consequences of determinacy hypotheses, each one proved in the year 1967.

5 Turing cones

The first example concerns the *degrees of unsolvability*, also called the *Turing degrees.* Each sequence of natural numbers, that is, each function $f : \mathbb{N} \to \mathbb{N}$, has a Turing degree $\mathbf{deg}(f)$, which should be thought of as a measure of the information that the f encodes. If f and g are functions, then $\mathbf{deg}(f) \leq \mathbf{deg}(g)$ if g has enough information to compute f mechanically, to answer mechanically questions about values of f. (The formal definition of $\leq$ is given in terms of the technical notion of *recursive in.*) Two different functions can have the same degree, as for instance do f and g if $g(n) = f(n) + 1$ for all n. The recursive functions all have the smallest degree, $\mathbf{0}$, because they are all computable by Turing machines. Two degrees can be incomparable. For example, two functions chosen in a sufficiently random fashion have incomparable degrees. For any two functions f and g, there is a least upper bound of $\mathbf{deg}(f)$ and $\mathbf{deg}(g)$, namely $\mathbf{deg}(h)$, where $h(2n) = f(n)$ and $h(2n+1) = g(n)$ for each n. The function h is the *join* of f and g.

If $\mathbf{d}$ is a degree, the *cone* of $\mathbf{d}$ is the set of all $\mathbf{d}'$ such that $\mathbf{d} \leq \mathbf{d}'$. A set of degrees is a *cone* if it is the cone of some degree.

The chapter (Slaman 1998), in this volume, discusses the Turing degrees and also discusses issues related to the lemma we are about to state.

One of the early consequences of the Axiom of Determinacy is the following, from (Martin 1968).

Cone Lemma *Assume that* AD *holds. If* $\mathbf{A}$ *is a set of Turing degrees, then either* $\mathbf{A}$ *contains a cone or else the complement of* $\mathbf{A}$ *contains a cone.*

Proof. Let $\mathbf{A}$ be a set of degrees. Let A be the set of all f such that $\mathbf{deg}(f)$ belongs to $\mathbf{A}$. By AD, either player I or player II has a winning strategy for the game G_A.

Suppose that I has a winning strategy σ. Now σ itself is not a function from $\mathbb{N}$ to $\mathbb{N}$, but it is easy to define a function $d : \mathbb{N} \to \mathbb{N}$ that encodes σ. Let $\mathbf{d} = \mathbf{deg}(d)$. We will prove that the cone of $\mathbf{d}$ is contained in $\mathbf{A}$. Let g be any function such that $\mathbf{deg}(g) \geq \mathbf{d}$. Let h be the play of the game produced when II 'plays g', that is, chooses $h(2n+1) = g(n)$ for each n, and I plays according to σ. Then h is the join of f and g, where $f(n) = h(2n)$ for each n. Since $\mathbf{deg}(g) \geq \mathbf{d}$, I's moves can be computed using g. In other words, $\mathbf{deg}(f) \leq \mathbf{deg}(g)$. Thus $\mathbf{deg}(h) \leq \mathbf{deg}(g)$. But also $\mathbf{deg}(g) \leq \mathbf{deg}(h)$. Thus $\mathbf{deg}(h) = \mathbf{deg}(g)$. Since σ is a winning strategy, $\mathbf{deg}(h) \in \mathbf{A}$, and so $\mathbf{deg}(g) \in \mathbf{A}$.

If II has a winning strategy for G_A, then a similar argument shows that the complement of $\mathbf{A}$ contains a cone. □

The proof just given is 'local', so that it gives consequences if one weakens the hypothesis AD to hypotheses consistent with the Axiom of Choice, that is, to *determinacy hypotheses*, as I am using the term. Say the set $\mathbf{A}$ of Turing degrees is *open*, *Borel*, *projective*, etc., if the set of all $f : \mathbb{N} \to \mathbb{N}$ such that $\mathbf{deg}(f) \in \mathbf{A}$ is open, Borel, projective, etc., respectively. By the proof of the Cone Lemma, Borel Determinacy implies that, if $\mathbf{A}$ is any Borel set of degrees, then either $\mathbf{A}$ or its complement contains a cone. Similarly, Projective Determinacy implies that every projective set of degrees or its complement contains a cone, etc. Since Borel Determinacy is provable in ZFC, the fact that every Borel set of degrees or its complement contains a cone is provable in ZFC. It is worth noting that the proof of Borel Determinacy came seven years after the proof of the Cone Lemma.

When I discovered the Cone Lemma, I became very excited. I was certain that I was about to achieve some notoriety within set theory by deducing a contradiction from AD. In fact I was pretty sure of refuting Borel Determinacy. I had spent the preceding five years as a recursion theorist, and I knew many sets of degrees. I started checking them out, confident that one of them would be give me my contradiction. But this did not happen. For each set I considered, it was not hard to prove, from the standard ZFC axioms, that it or its complement contained a cone. (Of course, one can use Choice to construct a counter-example to the unrestricted version of the theorem.)

I take it to be intuitively clear that we have here an example of prediction and confirmation. What was predicted, moreover, was not just individual assertions. Though there had been much work on the structure of the degrees, no attention at all had been paid to the notion of a cone. There was one known theorem (Richard Friedberg's 'criterion of completeness'), which we would now describe as showing that a certain set contains a cone. Afterwards cones and calculations of 'vertices' of cones became significant in degree theory. In determinacy theory, the Cone Lemma became an important tool. What was predicted by the Cone Lemma was thus a whole phenomenon, not merely isolated facts. The example seems fully analogous to striking instances of prediction and confirmation in empirical sciences.

6 Wadge degrees

I now turn to the second example: it is similar to the cone theorem but perhaps it is more basic and important.

Let $\mathbb{N}^{\mathbb{N}}$ be the set of all infinite sequences of natural numbers, that is, of all functions $f : \mathbb{N} \to \mathbb{N}$. A function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ is *continuous* if the value of the function $F(f)$ is determined by finitely many values of the function f (in the jargon of sequences, if each term of the sequence $F(f)$ is determined by finitely many terms of the sequence f). Under the natural topology giving rise to this notion of continuity (the product of the discrete topology on $\mathbb{N}$), $\mathbb{N}^{\mathbb{N}}$ is homeomorphic to the space of the irrational numbers with the usual topology. Thus we can almost, but not quite, think of our continuous functions as continuous functions from the reals to the reals.

Suppose that $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ is continuous and $B \subseteq \mathbb{N}^{\mathbb{N}}$. The *preimage* of B under F is the set A of all $f \in \mathbb{N}^{\mathbb{N}}$ such that $F(f) \in B$. For any topological notion of simplicity, the preimage A is at least as simple as B. If B is open, then A is open; if B is Borel, then A is Borel; etc. If one thinks of continuous functions as *computable*, then A is computable using B in the sense that a question of the form "Is $f \in A$?" can be answered by computing $F(f)$ and using an (oracular) answer to the question "Is $F(f) \in B$?".

These considerations motivate the notion of what are called 'Wadge degrees'. If A and B are subsets of $\mathbb{N}^{\mathbb{N}}$, then we say that the *Wadge degree* of A is $\leq$ the *Wadge degree* of B if there is a continuous $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ such that A is the preimage of B under F. The following result was proved by William Wadge in 1967.

Wadge's Lemma *Assume that* AD *holds. If A and B are subsets of* $\mathbb{N}^{\mathbb{N}}$, *then either the Wadge degree of A is $\leq$ that of B or else the Wadge degree of the complement of B is $\leq$ the Wadge degree of A.*

Proof Let A and B be subsets of $\mathbb{N}^{\mathbb{N}}$. We define a set $C \subseteq \mathbb{N}^{\mathbb{N}}$ by setting

$$\langle k_0, k_1, k_2, \ldots \rangle \notin C \;\leftrightarrow\; (\langle k_0, k_2, k_4, \ldots \rangle \in A \;\leftrightarrow\; \langle k_1, k_3, k_5, \ldots \rangle \in B).$$

The axiom AD implies that one of the players I and II has a winning strategy for the game G_C.

Assume first that τ is a winning strategy for II. The strategy τ defines a function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ as follows. For any sequence f, consider what happens when I plays the sequence f as $\langle k_0, k_2, \ldots \rangle$. Let $F(f)$ be the sequence $\langle k_1, k_3, \ldots \rangle$ that II, using τ, plays. The function F is continuous, for the n^{th} term of $F(f)$ depends only on the first n terms of f. Since τ is a winning strategy for II, it follows that A is the preimage of B under F. Thus the Wadge degree of A is $\leq$ that of B.

The existence of a winning strategy for I similarly implies that the Wadge degree of the complement of B is $\leq$ that of A. □

Like the proof of the Cone Lemma, the proof of Wadge's Lemma still gives results if we weaken the AD hypothesis to hypotheses consistent with Choice. Borel Deteminacy implies that the conclusion of Wadge's lemma holds for all Borel sets; Projective Determinacy implies that it holds for all projective sets; etc. Since Borel Determinacy is provable in ZFC, this latter consequence is provable in ZFC as well. As with the Cone Lemma, there was a seven-year gap between the proof of Wadge's Lemma and the proof of Borel Determinacy.

Remark. The proof of Borel Determinacy, though a proof in ZFC, is in a certain sense not an elementary proof. It uses the existence of uncountably many cardinal numbers. By the work of (Friedman 1971), any proof of Borel Determinacy must make some such appeal to uncountably many cardinal numbers. There is also an elementary proof of the Borel case of Wadge's Lemma, a proof not going through Borel Determinacy. This proof is due to Louveau and Saint-Raymond (1987 and 1988), and is much longer and more complex than the combination of the proofs of Borel Determinacy and Wadge's Lemma.

It follows directly from Wadge's lemma that, under AD, the Wadge degrees are linearly ordered (i.e., of any two, one is $\leq$ the other), except that a set and its complement may have incomparable Wadge degrees. (This is dramatically different from what happens with the Turing degrees.) Thus it is provable in ZFC that Wadge's ordering restricted to Borel sets is essentially linear, and determinacy hypotheses imply that this linearity holds for wider classes of sets.

It was later shown (by me, using an idea of Leonard Monk) that the Wadge linear ordering is actually a well-ordering. The proof, which may be found on pages 158–9 of (Kechris 1995), is not as simple as that of Wadge's Lemma, but it is nevertheless fairly short.

Remark. Everything we have said about Wadge degrees for subsets of $\mathbb{N}^{\mathbb{N}}$ goes through for subsets of $\mathbb{R}$, provided that the definition of $\leq$ is modified by an appropriate weakening of the continuity requirement.

Though the significance of continuous preimages has of course been well-known for a long time, the discovery of Wadge's lemma uncovered a phenomenon that is very basic in nature but had never been suspected: that all sufficiently simple sets (sets of reals, subsets of $\mathbb{N}^{\mathbb{N}}$, etc.) are arranged in a very natural well-ordered hierarchy of increasing complexity. There is no incomparability, except for the trivial kind mentioned above. For simple enough sets, namely Borel sets, the existence of the Wadge hierarchy is provable in ZFC. Determinacy hypotheses like Projective Determinacy imply that the hierarchy extends through wider classes of sets.

For many special cases of Wadge's Lemma, its conclusion has been verified by deduction from the ZFC axioms. A typical kind of instance is the following. When two naturally occurring sets are shown to have the same position in the Borel, projective, or another standard hierarchy, it can usually be shown that they are continuous preimages of one another, though the proof of this may be difficult. There were many instances of this phenomenon before Wadge's proof, and there have been many instances afterward. Of course, the proof of Borel Determinacy verified all Borel instances of Wadge's Lemma.

The conclusion of Wadge's Lemma is a powerful tool, useful in many contexts. Wadge himself used it to classify completely all the Borel sets.

7 Discussion

How far do the two examples I have described go toward establishing the truth of determinacy hypotheses like Projective Determinacy and $\mathrm{AD}^{L(\mathbb{R})}$? More importantly, could more—perhaps much more—evidence of the sort provided by the examples be sufficient for justifying scientific acceptance of determinacy hypotheses? (There is, in fact, a great deal of other evidence of this sort; but it is not my aim here to present the case for determinacy hypotheses, so I will not present this evidence.)

There is an oft-quoted passage from (Gödel 1947, page 477 of the expanded version) that I cannot resist quoting now:

> There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole discipline, and furnishing such powerful methods for solving problems (and even solving them, as far as that is possible, in a constructive way) that quite irrespective of their intrinsic necessity they would have to be assumed in the same way as any well-established physical theory.

My two examples exhibit for determinacy hypotheses all three of the properties Gödel mentions: abundant verifiable consequences; shedding light on a discipline; powerful methods for solving problems. (I do not understand Gödel's cryptic parenthetical remark well enough to know whether it applies to the examples.) We can certainly imagine a situation in which there was an enormous amount of evidence of the same general character for determinacy hypotheses. It seems clear that Gödel's conditions would then be satisfied. It also seems right that the hypotheses would have to 'be assumed in the same way as any well-established physical theory'.

In our imagined situation, would determinacy hypotheses count as *mathematically* known? In the jargon of §3 above, is the evidence of my examples mathematically proper evidence?

To ask this question is almost the same as asking whether, in the imagined situation, determinacy hypotheses would or should become basic axioms of mathematics. I will treat the two questions as identical, although this is not quite the case. One might, for example, deem the subject matter of determinacy hypotheses too specialized for them to be axioms of set theory. Perhaps one could do this while at the same time counting the hypotheses as mathematically known truths (and known because of the supposed evidence, not because, say, they followed from new axioms of a different kind).

If the answer to our question is 'yes', then—although the business of mathematics may be proving theorems—it is nevertheless not ruled out that among the basic axioms to be assumed without proof there are some whose epistemic status is like that of a well-confirmed physical theory.

If the answer is 'no', then there is a science different from mathematics but having the same subject matter as mathematics, a science that—though it may contain no empirical elements—yet has in essential ways the epistemic structure of an empirical science.

What is the answer? If determinacy hypotheses were not known to be independent of the standard ZFC axioms, then I am confident that the answer that would actually be given would be 'no'. One might say that there was strong evidence for the truth of determinacy hypotheses, but that the mathematical significance of this evidence was merely to give hope that the hypotheses could be proved.

If the kind of evidence in my examples is not mathematically proper, then precisely how does it fail to be so? In §3, I considered two candidates for earmarks of mathematically proper evidence. One candidate that I did not reject was that of being known *a priori*, but the evidence of the examples is as *a priori* as any mathematical evidence. Another candidate was certainty. However, I rejected certainty as a necessary condition for proper evidence, rejecting it on the grounds that the condition of certainty is not met by the evidence for some of the standard ZFC axioms.

In her comprehensive account (Maddy 1988*a*; 1988*b*) of the evidence for the ZFC axioms and proposed new axioms of set theory, Maddy distinguishes *intrinsic* and *extrinsic* evidence. The former is initially (Maddy 1988*a*, p. 482) described in terms of obviousness and self-evidence and later in terms of 'intuitiveness' (Maddy 1988*b*, p. 758). The latter is described as 'pragmatic, heuristic' (Maddy 1988*a*, p. 482). Maddy also talks of a third way of justifying axioms: *rules of thumb*, 'vague inutitions about the nature of sets, intuitions too vague to be expressed directly as axioms, but which can be used in plausibility arguments for more precise statements' (Maddy 1988*a*, p. 484). Later, however, she classifies the evidence for individual rules of thumb as intrinsic, extrinsic, and other. Thus, if we forget about 'other', we may think of a simple division into intrinsic and extrinsic evidence.

My two examples (like typical experimental examples in empirical sciences) count for Maddy as extrinsic. Moreover they no doubt fail to show what Gödel in the quoted passage calls 'intrinsic necessity', so it is perhaps fair to count Gödel also as classifying them as 'extrinsic'.

We can try to use the intrinsic–extrinsic distinction to clarify the notion of mathematically proper evidence. Specifically, we can consider the idea that only intrinsic evidence is mathematically proper.

Recall the attempt, discussed earlier, to argue for the ZFC axioms from the iterative concept of set. Such arguments, even if not yielding certainty, do seem to provide intrinsic evidence for the axioms. Furthermore, arguments from the iterative concept are not limited to the ZFC axioms. Remember that, according to the iterative concept, sets are to be regarded as being formed in a transfinite sequence of stages and that the number of these stages is supposed to be 'absolutely infinite'. From this absolute infinity one 'derives' the related principles of *resemblance* (there should be pairs of stages that are alike in any given respect) and *reflection* (there should be stages that look like the whole universe of sets in any given respect). From the reflection principle come precise *reflection schemata* in the formal language of set theory. One such reflection schema is equivalent with Infinity plus Replacement in the presence of the other ZFC axioms. Other reflection schemata imply the existence of various kinds of large cardinals, e.g., inaccessible cardinals. (See (Tait 1998) for a discussion of some reflection schemata and a strategy for justifying them.) The existence of various kinds of large cardinals can also be deduced from assertions motivated by the principle of resemblance. As I said earlier, all the determinacy hypotheses mentioned in this chapter have been deduced from large cardinal axioms—propositions stating the existence of large cardinals of various kinds. If principles like reflection and resemblance yield intrinsic evidence for the formal schemata obtained from them, then perhaps we can regard as known or supported by intrinsic evidence not just the ZFC axioms but also many large cardinal axioms and, indirectly, determinacy hypotheses.

One problem with the idea just enunciated is that the large cardinal axioms from which determinacy can be deduced are very strong ones. At the level of the large cardinal hierarchy where these axioms are found, the thread connecting the intuitive principles of resemblance and reflection with the large cardinal axioms is very thin indeed. Thus these intuitive principles provide little intrinsic justification for such strong large cardinal axioms. Though there is evidence for them, it is in fact overwhelmingly extrinsic. One might conclude from this that there is a real difference between the epistemic status of weak and strong large cardinal axioms. Perhaps the standard ZFC axioms and weak large cardinal axioms (as in (Tait 1998)) have intrinsic—and so mathematically proper—evidence, while stronger large cardinal axioms and determinacy are only extrinsically supported. Perhaps, consequently, there is mathematically proper evidence for the former but not for the latter.

I do not think it implausible to regard, in the way just indicated, ZFC and weak large cardinal axioms as supported by a larger body of intrinsic evidence than that which supports strong large cardinal axioms and determinacy. But I do not think that this provides a significant difference between the total body of evidence for the two classes of propositions. The reason is that, even for some of the ZFC axioms, the intrinsic evidence is not the main evidence. Consider, for example, the Axiom of Infinity. There is certainly intrinsic evidence for it. But there is intrinsic evidence against it that is at least as compelling: the intuitive idea that there can be no completed infinite totalities, for example. The reason for Infinity's acceptance is not that the intrinsic evidence for it is stronger than that against. The reason, as Maddy indirectly says, is evidence that is 'predominantly extrinsic, lying in the depth, breadth and effectiveness of the subject it launched' (Maddy 1988*b*, p. 759), (that is, Cantorian set theory). Similar points can be made for other ZFC axioms and for weak large cardinal axioms such as that asserting the existence of an inaccessible cardinal. Woodin (1998) might be taken as making a similar point even for axioms of Peano Arithmetic, although he is talking of consistency rather than of truth.

I believe, then, that the attempt to make being intrinsic the earmark of mathematically proper evidence will not stand up. Like the corresponding attempt based on certainty, it fails because it does not explain the properness of the evidence for the standard ZFC axioms.

In general, I do not see any convincing rationale for ruling out as mathematically proper the kind of evidence provided by my two examples. Nor do I feel that such evidence's being mathematically proper is unsatisfying or counterintuitive. It does not follow that mathematics is indistinguishable from empirical sciences. One can still hold that mathematics is an *a priori* science. Moreover, the character of mathematics can still reside in proofs, whatever the grounds for accepting the basic axioms.

Bibliography

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*, **70**, 661–80. Reprinted in *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 272–94. Cambridge University Press, 1983.

Benacerraf, P. and Putnam, H. (ed.) (1983). *Philosophy of mathematics: selected readings* (2nd edn). Cambridge University Press.

Davis, M. (1964). Infinite games of perfect information. In *Advances in game theory* (ed. M. Dresher, L. S. Shapley, and A. W. Tucker), pp. 85–101. Annals of Mathematics Studies, Vol. 52. Princeton University Press.

Friedman, H. M. (1971). Higher set theory and mathematical practice. *Annals of Mathematical Logic*, **2**, 325–57.

Gale, D. and Stewart, F. M. (1953). Infinite games with perfect information. In *Contributions to the theory of games,* Vol. 2 (ed. H. Kuhn and A. Tucker), pp. 245–66. Annals of Mathematics Studies, Vol. 28. Princeton University Press.

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.

Kechris, A. S. (1995). *Classical descriptive set theory.* Springer-Verlag, Heidelberg.

Louveau, A. and Saint-Raymond, J. (1987). Borel classes and closed games: Wadge-type and Hurewicz-type results. *Trans. American Math. Soc.*, **304**, 431–67.

Louveau, A. and Saint-Raymond, J. (1988). The strength of Borel Wadge Determinacy. In *Cabal Seminar 1981–1985* (ed. A. S. Kechris, D. A. Martin, and J. R. Steel), Lecture Notes in Mathematics, Vol. 1333, pp. 1–30. Springer-Verlag, Berlin.

Maddy, P. (1988*a*). Believing the axioms. I. *Journal of Symbolic Logic*, **53**, 481–511.

Maddy, P. (1988*b*). Believing the axioms. II. *Journal of Symbolic Logic*, **53**, 736–64.

Martin, D. A. (1968). The axiom of determinateness and reduction principles in the analytical hierarchy. *Bull. American Math. Soc.*, **74**, 687–68.

Martin, D. A. (1975). Borel determinacy. *Annals of Mathematics*, **102**, 363–71.

Martin D. A. and Steel, J. R. (1989). A proof of projective determinacy. *Journal American Math. Soc.*, **2**, 71–125.

Mauldin, R. D. (ed.) (1981). *The Scottish book: mathematics from the Scottish café.* Birkhäuser, Boston.

Moore, G. H. (1982). *Zermelo's Axiom of Choice: its origins, development, and influence.* Springer-Verlag, New York.

Moschovakis, Y. N. (1980). *Descriptive set theory.* North-Holland, Amsterdam.

Mycielski, J. and Steinhaus, H. (1962). A mathematical axiom contradicting the axiom of choice. *Bulletin de l'Académie Polonaise des Sciences, Série des Sciences Mathématiques, Astronomiques et Physiques*, **10**, 1–3.

Slaman, T. A. (1998). Mathematical definability. *This volume*, 233–51.

Tait, W. (1998). Foundations of set theory. *This volume*, 273–90.

Woodin, W. H. (1988). Supercompact cardinals, sets of reals and weakly homogeneous trees. *Proceedings of the National Academy of Science USA*, **85**, 6587–91.

Woodin, W. H. (1998). The tower of Hanoi. *This volume*, 329–51.

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae*, **16**, 29–47.

Departments of Mathematics and Philosophy
University of California
Los Angeles
CA 90095-1555
USA
email: dam@math.ucla.edu

13

Mathematical definability

Theodore A. Slaman

1 Introduction

One might fairly say that the mathematical analysis of definability began in 1931, with the appearance of Gödel's Incompleteness Theorem (Gödel 1931). Gödel showed that, for sufficiently strong formal systems T, there exist undecidable statements φ such that there is no proof of φ or of $\neg\varphi$ within T. This theorem pointed to an intrinsic incompleteness within the formal notion of proof.

The method of computation by algorithm is more general than that of verification by formal proof. It too was shown to be incomplete, but it took some time to develop the technical apparatus needed to state this incompleteness correctly. (Kleene 1987), in his biographical memoir of Gödel, recalls this development and we summarize some of his remarks. Kleene describes the intuitive notion of an algorithm as follows.

> An *algorithm* is a procedure described in advance such that, whenever a value is chosen for the variable or a respective value for each of the variables of the function (or predicate), the procedure will apply and enable one in finitely many steps to find the corresponding value of the function (or to decide the truth or falsity of the corresponding value of the predicate).

Several specific implementations of the notion of algorithmic procedure were proposed; see (Church 1936*a*), (Gödel 1965), (Kleene 1936), (Post 1936), and (Turing 1936). Upon the discovery that, while the details of presentation varied from one to the other of these models of computation, the same class of functions (the recursive functions) was declared computable by all of them, the Church–Turing Thesis, that the collection of computable, or effectively calculable, functions is exactly the collection of recursive functions, was generally accepted.

Definition 1.1

- *A function or relation is* recursive *if and only it is computable on a Turing machine.*
- *A relation is* recursively enumerable *if and only if it is the domain of a recursive function.*

I was partially supported by NSF Grant DMS-9500878 during the preparation of this chapter. I wish to thank A. Kechris for suggesting corrections to a preliminary draft of this chapter.

- *A function or relation is* recursively approximated *if it is the pointwise limit of a uniformly recursive sequence. For example, if g is a recursive function of two variables such that* $\lim_{s\to\infty} g(x,s)$ *exists for all x, then this limit is recursively approximated.*

We define the relation X *is recursive relative to* Y similarly, and denote it by $Y \geq_T X$.

With the precise definition of a recursive function came the proofs of the existence of arithmetically definable sets which are not recursive. Kleene credits (Church 1936*b*) and (Turing 1936) with showing

> the existence of 'intuitively undecidable predicates', that is, predicates for which there is no 'decision procedure' or 'effective process' or 'algorithm' by which, for each choice of a value of its variable, we can decide whether the resulting proposition is true or false.

Kleene goes on to claim for himself the establishment of the connection between the proof theoretic and the computational lines of development. In short, in any reasonable proof system there is an algorithm to recognize the valid proofs. If T were to be a complete axiomatization of number theory, then the truth of any number theoretic statement φ would be determined recursively by finding either a proof of φ from T or a proof of $\neg\varphi$ from T. This would contradict the Church and Turing theorems.

Kleene summarizes his reaction to these incompleteness theorems by saying that the theory of the natural numbers 'offers inexhaustible scope for mathematical ingenuity'. No recursive axiom system is sufficient to prove all of the true number-theoretic statements. Similarly, no countable syntax and semantics are sufficient to describe all of the definable sets of natural numbers.

One way to gain mathematical leverage is to expand the metamathematical environment. (Gödel 1932) noted that if elementary number theory is

> ...successively enlarged by the introduction of variables for classes of numbers, classes of classes of numbers, and so forth, together with the corresponding comprehension axioms, we obtain a sequence (continuable into the transfinite) of formal systems ... it turns out that the consistency (ω-consistency) of any of those systems is provable in all subsequent systems.

By the same enlargement in language, we obtain a sequence of families of definable sets. Analogously, the universal set for any of those families is an element of all subsequent families. This is the starting point for the discussion below.

In §2 we will give an introduction to the hierarchy of definability in first- and second-order arithmetic. On the one hand, we ignore distinctions of finite size and thereby miss the details of computational complexity. On the other hand, we will give a short treatment of those aspects of definability which are tied to axiomatic set theory and the large cardinal hierarchy.

In §3 we will illustrate the utility of a detailed structure theory for definability. We give mathematical examples which exactly occupy distinguished positions within the hierarchy, such as arithmetically definable sets which are not recursive and analytically definable sets which are not Borel. We also discuss how insights into definability lead to insights into provability within second-order arithmetic. We end in §4 with a discussion of whether the proposed hierarchy of definability is intrinsic, with the conclusion that it is so.

2 The intuitive hierarchy

First-order arithmetic We generate the formulas in a first-order language by recursion from logical symbols (,), $\longrightarrow$, $\neg$, and $=$; variables $x_1, x_2, \ldots$; quantifiers $\exists$ and $\forall$; and non-logical symbols for constants, functions, and relations. The first-order language of arithmetic includes the constants 0 and 1, the functions $+$ and $\times$, and the binary relation $<$.

We use $(\forall x < y)\varphi$ or $(\exists x < y)\varphi$ as abbreviations for $(\forall x)[x < y \longrightarrow \varphi]$ and $(\exists x)[x < y \longrightarrow \varphi]$. We say that $(\exists x < y)$ and $(\forall x < y)$ are *bounded quantifiers.*

The standard hierarchy for definability within first-order arithmetic is based on counting alternations between unbounded quantifiers. One can also keep track of the bounded quantifiers, and be led to questions of computational complexity.

Definition 2.1 *Let φ be a formula in first-order arithmetic. Then:*

- *φ is Σ^0_0 and Π^0_0 if it has no quantifiers which are not bounded;*
- *φ is Σ^0_{n+1} if it is of the form $(\exists x_1) \cdots (\exists x_m)\psi$, where ψ is Π^0_n; φ is Π^0_{n+1} if it is of the form $(\forall x_1) \cdots (\forall x_m)\psi$, where ψ is Σ^0_n.*

For example, $(\exists x)(\forall y)(x + y = y)$ is a Σ^0_2 sentence.

We say that a predicate R on the natural numbers is Σ^0_n or Π^0_n if it has a Σ^0_n or Π^0_n definition. That is, there is a Σ^0_n or Π^0_n formula $\varphi(x_1, \ldots, x_m)$ such that, for all $n_1, \ldots, n_m$ from $\mathbb{N}$,

$$R(n_1, \ldots, n_m) \quad \text{if and only if} \quad \langle \mathbb{N}, 0, 1, +, \times, < \rangle \models \varphi(n_1, \ldots, n_m).$$

We say that R is Δ^0_n if it is both Σ^0_n and Π^0_n.

It is worth mentioning that bounded quantifiers do not amount to much in this hierarchy. If φ is a Σ^0_n formula, then the relation defined by $(\forall x < y)\varphi$ is also Σ^0_n. Similarly, if φ is a Π^0_n formula, then the relation defined by $(\exists x < y)\varphi$ is also Π^0_n.

The first levels of the syntactic hierarchy have dynamic interpretations.

Theorem 2.2 *Let R be a predicate on the natural numbers. Then:*

- *R is Δ^0_1 if and only if it is recursive;*
- *R is Σ^0_1 if and only if it is recursively enumerable;*
- *R is Δ^0_2 if and only if it is recursively approximated.*

The shortest route to a set of integers which is not recursive is through Cantor's theorem that the power set of the natural numbers is not countable. Since there are only countably many Turing machines, there is a set of integers which is not computed by any Turing machine. But this argument is not the one to which we referred above. We seek a set which is definable, even definable in elementary number theory, and which is not recursive.

There is a canonical way to produce a definable set which is not recursive, the *diagonal method.*

- Show that there is a recursive way to use single numbers to represent finite sequences of numbers.
- Show that there is a *universal* Σ^0_1 predicate: First, fix a recursive enumeration of all of the Σ^0_1 formulas of one free variable, $\langle\varphi_e(y) : e \in \mathbb{N}\rangle$. Then, show that there is a Σ^0_1 formula $\psi(e, y)$ such that for all e and y, $\psi(e, y)$ is equivalent to $\varphi_e(y)$.
- If $\psi(e, y)$ did define a recursive set of pairs, then there would be an e_0 such that, for all e, $\neg\psi(e, e)$ if and only if $\varphi_{e_0}(e)$. Evaluate this equivalence at e_0 to get $\neg\psi(e_0, e_0)$ if and only if $\varphi_{e_0}(e_0)$, which contradicts the equivalence between $\psi(e_0, y)$ and $\varphi_{e_0}(y)$.

The above is a purely syntactic argument. It does not use any detailed information about definability in first-order arithmetic, and so it applies in a wide variety of situations. It points out that whenever there is a universal predicate within some definability class, then negating elements of that class results in greater expressive power.

To be precise, we define the existence of a universal set as follows.

Definition 2.3 *Suppose that Γ is a collection of sets, and let $\Gamma(X)$ be those elements of Γ which are subsets of X. Then a subset U of $(\mathbb{N}\times X)\cap\Gamma$ is* universal *for $\Gamma(X)$ if $\Gamma(X)$ is equal to the collection of sets $U_e = \{x : U(e, x)\}$.*

We will speak simply of the existence of a universal set when there is little risk of confusion about the intended values of X. The special case of the universal Σ^0_1 set will be important.

Definition 2.4 *For each A, a subset of $\mathbb{N}$ or function from $\mathbb{N}$ to $\mathbb{N}$, we let A' denote the universal Σ^0_1 set relative to A. The set A' is called the* Turing jump *of A.*

The diagonal argument can be used to conclude the following.

Theorem 2.5 *For each n, there is a universal Σ^0_n relation and it is not Π^0_n.*

Similarly, for each n, there is a Δ^0_{n+1} relation which is not Σ^0_n and not Π^0_n. The set of sentences in the first-order language of arithmetic which are true of the natural numbers is not Σ^0_n, for any n.

This diagonal argument may apply in wide generality. However, even in its original form it does not show that any particularly interesting set is not recursive, but only that some recursively enumerable set is not recursive.

The fixed point theorem The diagonal argument given above implies that there is no universal total recursive function. By this we mean that there is no recursive function f such that for all e and m, $f(e, m)$ is defined and for which, for every total recursive g, there is an e such that for all m, $g(m) = f(e, m)$. If f

were to be total recursive, then the diagonal function $g(m) = 1 + f(m, m)$ would also be recursive, and this conclusion would lead to a contradiction.

We recall the Kleene Fixed Point Theorem. We use Kleene's notation and let $\{e\}$ denote the e^{th} recursive function, in the standard enumeration of recursive functions by the programs which compute them.

Theorem 2.6 (Kleene) *For any recursive function f there is an e such that $\{e\} = \{f(e)\}$.*

One interpretation of Theorem 2.6 is that the class of recursive functions is impervious to attempts to build a function which is not recursive by means of an effective diagonal argument. Shoenfield (1995) remarks that it was Theorem 2.6 that convinced Kleene of the validity of the Church–Turing hypothesis.

If we drop the constraint that f must be defined at all arguments, so we include computations that do not halt, then the existence of a universal Σ^0_1 predicate is essentially the same as the existence of a universal partial recursive function.

Second-order arithmetic We obtain the language of second-order arithmetic by adding new function variables $F_1, F_2, \ldots$. We allow for forming a new term $F_i(t)$ when t is a term and allow quantifiers to range over number theoretic and function variables.

It is typical in mathematical logic to use *analysis* and *second-order arithmetic* synonymously and to refer to functions on the natural numbers as *reals.* What is really meant here is that we work within a mathematical realm in which everything is countable, except that the quantifiers range over an uncountable collection of countable objects.

We define the (lightface) *projective hierarchy* as follows.

Definition 2.7 *Let φ be a formula in second-order arithmetic. Then:*

- *φ is Σ^1_0 and Π^1_0 if it has no quantifiers over real variables;*
- *φ is Σ^1_{n+1} if it is of the form $(\exists F_1) \cdots (\exists F_m)\psi$, where ψ is Π^1_n;*
- *φ is Π^1_{n+1} if it is of the form $(\forall F_1) \cdots (\forall F_m)\psi$, where ψ is Σ^1_n.*

As in first-order arithmetic, R is Σ^1_n or Π^1_n if it has a Σ^1_n or Π^1_n definition and R is Δ^1_n if it is both Σ^1_n and Π^1_n.

Just as bounded quantifiers (quantifiers over finite sets) do not contribute within the arithmetic hierarchy, number theoretic quantifiers do not contribute in the projective hierarchy. Let x be a number theoretic variable. If φ is a Σ^1_n formula, then the relation defined by $(\forall x)\varphi$ is also Σ^1_n. Similarly, if φ is a Π^1_n formula, then the relation defined by $(\exists x)\varphi$ is Π^1_n.

There is an obvious and profound difference between definability within first- and second-order arithmetic. Every natural number is definable within first-order arithmetic: 0 is explicitly defined by a constant symbol and each $n > 0$ is defined by the term obtained by adding 1 to itself n times. Consequently, if R is defined by a formula with a natural number parameter, then there is a definition of R by a formula of the same arithmetic complexity in which there are no parameters. But there are uncountably many reals, only countably many of which are definable. Thus, more sets become definable when one uses parameters. To take a trivial example, for each real A, $\{A\}$ is definable using A as a parameter.

The collection of Borel sets of reals makes a more substantial example. We can generate the Borel sets by transfinite recursion through the countable ordinals: start with the open sets and iterate the operations of complementation and countable union. The same collection is obtained by starting with the open intervals with rational endpoints, each of which is definable without parameters. However, an arbitrary countable union has the same complexity as an arbitrary real number, and so real number parameters appear in an essential way in the Borel hierarchy.

Definition 2.8 *A relation R is Σ^1_n if there are real parameters $P_1, \dots, P_n$ and a Σ^1_n formula $\varphi(F_1, \dots, F_m, F_{m+1}, \dots, F_{m+n})$ such that R is the set of solutions to $\varphi(F_1, \dots, F_m, P_1, \dots, P_n)$. Define Π^1_n, Δ^1_n, Σ^0_n, Π^0_n, and Δ^0_n, similarly.*

The first levels of the syntactic hierarchy can be formulated in classical descriptive set theoretic terms. The *projective sets* are formed by starting with the Borel sets, in arbitrary finite dimension, and closing under complementation and projection. At the first level, the *analytic sets* are the projections of Borel sets and the *co-analytic sets* are their complements.

Theorem 2.9 *For any predicate R:*

- *R is Δ^1_1 if and only if it is Borel;*
- *R is Σ^1_1 if and only if it is analytic;*
- *R is Π^1_1 if and only if it is co-analytic;*
- *R is projective if and only if there is an n such that R is Σ^1_n.*

The specific interest in definability lies in understanding those mathematical objects specified without parameters. We discuss a few examples.

The Δ^1_1 sets The syntactic presentation of the Δ^1_1 sets is by pairs of complementary Σ^1_1 predicates. (Kleene 1955) shows that there is a transfinite recursion to generate exactly the Δ^1_1 sets: essentially, they are the effectively presented Borel sets. Kleene introduces the following effective theory of the countable ordinal numbers and of the Borel operations to generate the Δ^1_1 sets.

- Define a system of notations $\mathcal{O}$ for ordinals with recursive operations of successor, limit, ordinal addition and so forth. Use the Kleene Fixed Point Theorem to define recursive functions on $\mathcal{O}$ by transfinite recursion.
- Associate a real number H_a with each notation a in $\mathcal{O}$ so that H_0 is the empty set, ordinal succession corresponds to the Turing jump, and countable limits correspond to uniform amalgamation.
- Say a real is *hyperarithmetic* if it is recursive relative to some H_a. Similarly, define a real to be hyperarithmetic in F by starting with $H^F_0 = F$.
- Define a set of reals R to be hyperarithmetic if there is an a and a recursive function f such that the following conditions hold for all F: $a \in \mathcal{O}^F$, and $F \in R$ if and only if $f(H^F_a) = 0$.

Theorem 2.10 (Kleene) *A real or a set of reals is Δ^1_1 if and only if it is hyperarithmetic.*

The Π^1_1 sets In Kleene's analysis of the Δ^1_1 sets, he not only provided a detailed structure theory for the Δ^1_1 sets but also for the Π^1_1 sets.

First, reduce to the following normal form for Π^1_1 predicates. Using the effective pairing function and absorbing quantifiers over natural number variables, for each Π^1_1 set R there is a Σ^0_1 predicate $(\exists n)\varphi(n, F)$ such that R is defined by the formula $(\forall F)(\exists n)\varphi(n, F)$. Here, we mean that φ may have other free variables but no other quantifiers other than bounded number quantifiers.

Second, given a formula $(\forall F)(\exists n)\varphi(n, F)$, define the tree T of functions F_0 with some finite domain $[0, \dots, m]$ such that $\neg(\exists n \leq m)\varphi(n, F_0)$. In other words, T is the tree of possible initial segments of counter-examples to $(\forall F)(\exists n)\varphi(n, F)$. Then $(\forall F)(\exists n)\varphi(n, F)$ is equivalent to the statement that T is wellfounded (that is, has no infinite path). So the set

$$\{e : \text{the } e^{\text{th}} \text{ recursive binary relation is a wellfounded tree.}\}$$

is a universal Π^1_1 set of natural numbers. Similarly, the set

$$\{e : \text{the } e^{\text{th}} \text{ recursive in } F \text{ binary relation is a wellfounded tree.}\}$$

is a universal Π^1_1 set of reals.

The elements of a Δ^1_1 set are determined by means of a transfinite recursion, whose length can be specified in advance. The elements of a Π^1_1 set are also determined by a transfinite recursion, but it only converges on the elements of the set, and diverges on the elements of the complement. In this way, the Δ^1_1 sets and the Π^1_1 sets are analogous to the recursive and recursively enumerable sets, respectively.

The Π^1_2 sets When we consider Π^1_2 predicates, we must also consider metamathematical questions. We begin with Shoenfield (1961) Absoluteness Theorem. Here ZF denotes the Zermelo–Fræenkel axioms of set theory.

Theorem 2.11 (Shoenfield) *Suppose that $\mathcal{M}_1$ and $\mathcal{M}_2$ are models of* ZF, *$\mathcal{M}_1$ is a submodel of $\mathcal{M}_2$ containing all of the ordinals of $\mathcal{M}_2$ and φ is a Π^1_2 predicate with parameters from $\mathcal{M}_1$. Then, $\mathcal{M}_1 \models \varphi$ if and only if $\mathcal{M}_2 \models \varphi$.*

We say that Π^1_2 statements are *absolute* between wellfounded inner models of ZF.

Gödel's universe of constructible sets Gödel (1938) proved the consistency of the Continuum Hypothesis and of the Axiom of Choice by a remarkable and beautiful analysis of definability. We mention a few of the highlights of Gödel's proof and its consequences for the Π^1_2 sets. (In the following, we will work within a model of set theory and prove consistency, rather than work within arithmetic and prove relative consistency, but the same proof can be used for both purposes.)

Define a hierarchy of sets by:

$$\begin{aligned} L_0 &= \emptyset; \\ L_{\beta+1} &= \{x : x \text{ is first-order definable in parameters over } L_\beta\}; \\ L_\lambda &= \textstyle\bigcup_{\beta<\lambda} L_\beta . \end{aligned}$$

Say x is *constructible* if and only if there is an α such that x is an element of L_α. Let L be the class of constructible sets.

Showing that L is a model of the axioms of set theory involves a reflection argument. Suppose that φ is a formula using parameters from L. Then, for arbitrarily large ordinals α, for all x in L_α,

$$L_\alpha \models \varphi(x) \quad \text{if and only if} \quad L \models \varphi(x).$$

Thus the closure properties of the class of ordinals imply that closing under local definability at each step α implies closure under global definability in the limit.

The constructible sets are presented with an intrinsic definable well-ordering coming from the order of their construction. Thus, L is a model which establishes the consistency of the Axiom of Choice.

Gödel's proof that L is a model of the Continuum Hypothesis requires a further insight. Suppose that X is a constructible real number. Then there is an ordinal α such that X is an element of L_α. Let $\mathcal{H}$ be a countable elementary substructure of L_α with $X \in \mathcal{H}$, and let $\pi : \mathcal{H} \to \mathcal{H}^*$ be its transitive collapse. Then π is the identity on $\mathbb{N}$, and so X is an element of $\mathcal{H}^*$. Now, $\mathcal{H}^*$ is a transitive structure, built by transfinitely iterating first-order definability over the empty set. Thus, $\mathcal{H}^*$ is a countable initial segment of L, say $\mathcal{H}^* = L_{\alpha^*}$. Then X is an element of $L_{\omega_1^L}$, where ω_1^L is the least ordinal which is not countable in L. There is a constructible bijection between $L_{\omega_1^L}$ and ω_1^L, so L satisfies the statement that there are at most ω_1 reals, and thereby satisfies the Continuum Hypothesis.

Gödel could also draw some further information from the above argument. Given two reals X and Y, X is constructible before Y if and only if

$$(\exists \alpha < \omega_1^L)[L_\alpha \models X \text{ is constructible before } Y].$$

We can rewrite this condition as

$$(\exists Z)\left[\begin{array}{c} Z \text{ codes a countable initial segment of } L \\ \text{and } X \text{ is constructible before } Y \text{ in the coded model.}\end{array}\right]$$

A countable model $\mathcal{M}$ is isomorphic to an initial segment of L if and only if $\mathcal{M}$ is wellfounded and built by transfinitely iterating first-order definability over the empty set. The latter condition is Π^1_1. Consequently, within L there is a Σ^1_2 well-ordering of the reals. Of course, once one has a definable well-ordering of the reals, then the various pathologies associated with the Axiom of Choice are realized by definable sets of reals. For example, in L there is a Σ^1_2 set which is not Lebesgue measurable.

On the other hand, we can use L to mathematical advantage through the

Shoenfield Absoluteness Theorem. Suppose that φ is a Π^1_2 statement. Then φ is true if and only if φ is satisfied within L. Consequently, if we can prove φ using the Axiom of Choice, the Continuum Hypothesis or any of the other structure theory of L, then we may conclude that φ is a theorem of ZF. Kechris (1991) includes a very nice example of an argument of this type.

The Δ^1_3 sets By the previous discussion, 'X is constructible' is a Σ^1_2 statement about X. Consequently, 'There is a nonconstructible real' is a Σ^1_3 sentence. Simultaneously with proving the consistency of the failure of the Continuum Hypothesis, Cohen (1966) proved that it is consistent (with ZFC) that there is a real number which is not constructible. Cohen introduced the method of *forcing* to add new sets to a given (countable) model $\mathcal{M}$ of ZF.

Briefly, the new sets are approximated by *conditions* in $\mathcal{M}$. For example, Cohen could add a new subset of ω using the collection of finite approximations to such a set as the conditions. The conditions are partially ordered so that stronger conditions give more information about the set being approximated. Then one considers only generic sets, that is sets which realize every possible behavior as measured by $\mathcal{M}$. Precisely, one considers sets which meet every dense subset $D \in \mathcal{M}$ of the forcing partial order. The model $\mathcal{M}[G]$ is obtained starting from a generic set G and the elements of $\mathcal{M}$, and transfinitely iterating first-order definability through the ordinals of $\mathcal{M}$. $\mathcal{M}[G]$ is well approximated within $\mathcal{M}$ by what is called the *forcing relation.* The closure of $\mathcal{M}$ implied by its being a model of ZF implies a similar closure of $\mathcal{M}[G]$, and so $\mathcal{M}[G]$ is also a model of ZF.

It might seem that these sets produced by forcing, because of their generic nature, could not be individually definable. Kreisel raised the following question. Does every set of natural numbers which is Σ^1_n, for some n, belong to L? An affirmative answer would have made for a very simple analysis of the definable reals but such is not the case. In fact, the simplest possible counterexamples are possible and can even be found by forcing.

Theorem 2.12 (Jensen and Solovay 1970) *If* ZFC *is consistent, then so is* ZFC *with the statement:*

There is a Δ^1_3 real number which is not constructible.

Large cardinals The most remarkable, or perhaps the most natural, development in the theory of definability within second-order arithmetic is its dependence on metamathematical considerations and especially its interactions with the global properties of the universe of sets. By the Lowenheim–Skolem Theorem, one should anticipate large scale phenomena to reflect to countable ones. But the opposite occurs as well. Unfortunately, we cannot do justice to the beautiful mathematics in this area, but we will mention some of its first generation of results.

Definition 2.13 *A set of* order indiscernibles *for a model $\mathcal{M}$ is a linearly ordered set I from $\mathcal{M}$ such that for any $a_1 < \cdots < a_n$ and $b_1 < \cdots < b_n$ and* *any formula in the language of* $\mathcal{M}$,

$$\mathcal{M} \models \varphi(a_1, \ldots, a_n) \text{ if and only if } \mathcal{M} \models \varphi(b_1, \ldots, b_n).$$

Gaifman (1964) shows that the existence of a measurable cardinal implies that there is a unique class of ordinals S containing all uncountable cardinals such that, for all uncountable cardinals κ,

- $S \cap \kappa$ has order type κ, and if κ is regular, then $S \cap \kappa$ is a closed and unbounded subset of κ;
- $S \cap \kappa$ is a set of order indiscernibles for $\langle L_\kappa, \epsilon \rangle$;
- every element of L_κ is definable within L_κ using parameters from $S \cap \kappa$.

Such a set S is called a set of *Silver indiscernibles* for L. It follows from the above properties that, for each α in S, L_α is an elementary substructure of L. We define $0^\#$ to be (the codes for) the elementary theory of L_{ω_1} relative to parameters from S. We write 'exists $0^\#$' to indicate that there is such a class of indiscernibles for L.

To give an idea of its metamathematical strength, we show that if $0^\#$ exists, then every definable element of L is countable, though not necessarily countable in L. Suppose that a is the unique solution to φ in L; then $L \models (\exists x)(\varphi(x))$ and so $L_{\omega_1} \models (\exists x)\varphi(x)$; every element of L_{ω_1} is countable; hence, a is countable.

Now, we go to the connection with definability in second-order arithmetic. Solovay showed that, if $0^\#$ exists, then it is a Δ^1_3 real, and that there is a Π^1_2 formula θ such that (provably within ZFC) θ has at most one solution, which is $0^\#$, if it exists.

3 Applications

Matijasevič's theorem Matijasevič's solution (Matijasevič 1970) to Hilbert's tenth problem (Hilbert 1901–2) is a canonical example of an undecidability theorem.

Hilbert's tenth problem is equivalent to the following one: find an algorithm to determine whether a given polynomial with integer coefficients has a solution within the natural numbers. That is to say that Hilbert's problem is to give an algorithm to tell whether a given Diophantine equation has a solution within the natural numbers.

Matijasevič showed that every set of natural numbers which is defined by a Σ^0_1 formula is also defined as being the set of solutions within the natural numbers of some Diophantine equation. Of course, this part of the proof involves a substantial analysis of Diophantine equations. It follows from the fact that the universal Σ^0_1 subset of $\mathbb{N}$ is not recursive that there is no recursive algorithm as required by Hilbert.

We can also use Matijasevič's theorem to refine our discussion of the arithmetic hierarchy. We defined a Σ^0_1 formula to be one of the form $(\exists x_1) \cdots (\exists x_n)\varphi$, where φ had no unbounded quantifiers. We explicitly allowed φ to have bounded quantifiers. Now, by Matijasevič's theorem, we know that every Σ^0_1 set is defined by a Diophantine equation and so using the bounded quantifiers did not result in any new sets being defined.

Of course, there is a large number of undecidability results, sufficiently well known that we will not give further examples here. In fact, nontrivial decidability theorems are harder to find than undecidability ones.

Compactness Typically, a compactness theorem asserts the existence of an infinite set under a finite closure hypothesis. For example, König's Lemma states that, if T is a binary tree with arbitrarily large finite branches, then T has an infinite branch. Similarly, the compactness theorem of first-order logic states that any finitely satisfiable first-order theory is satisfiable. When we apply these compactness theorems, we construct a particular tree and are willing to accept any path through it, or we construct a consistent theory and are willing to accept any model in which it is satisfied. We must be content to work with a generic path or a generic model.

We look closely at some examples and draw some conclusions concerning Friedman's (1975) subsystems of second-order arithmetic. Define RCA_0 to be the axiom scheme specifying that the natural numbers satisfy induction for Σ^0_1 formulas and that the reals are closed under relative computability.

Definition 3.1 *A* Turing ideal *is a nonempty set of reals $\mathcal{T}$ such that, for all $X \in \mathcal{T}$ and all Y, if $X \geq_T Y$, then $Y \in \mathcal{T}$.*

So RCA_0 is something of an axiomatization of the theory of a Turing ideal.

Friedman compares systems by their sets of theorems. We will pay attention to the complexities of their first differences.

Definition 3.2 *Suppose that T_1 and T_2 are first-order theories and Γ is a set of sentences in their common language. We say that T_2 is* Γ-conservative over *T_1 if, for all $\varphi \in \Gamma$, if φ is a theorem of T_2, then φ is a theorem of T_1.*

König's lemma

Definition 3.3 *A* Scott set *is a Turing ideal $\mathcal{S}$ such that if $X \in \mathcal{S}$ and T is a binary tree recursive in X, then T has an infinite path in $\mathcal{S}$.*

Questions about Scott sets measure what can be built (recursively) from applications of König's lemma or equivalently from applications of the compactness theorem in first-order logic.

Theorem 3.4 (Jockusch and Soare 1972)

- *Every binary infinite tree T has an infinite path X such that $X' \leq_T T'$.*
- *There is a Scott set $\mathcal{S}$ such that for all $X \in \mathcal{S}$, $X' \leq 0'$.*

One interpretation of Theorem 3.4 is that applications of compactness do not generate complicated sets. Let *WKL* be the statement of König's lemma for binary trees given above. Harrington, in an unpublished result, used the Jockusch and Soare machinery to prove the following result.

Theorem 3.5 (Harrington 1978) *$RCA_0 + WKL$ is Π^1_1-conservative over RCA_0.*

In fact, any statement of elementary number theory which is proven by $RCA_0 + WKL$ can be proven in first-order arithmetic using only induction for Σ^0_1 formulas, without mention of infinite sets or of compactness.

Ramsey's theorem

Definition 3.6 *For $X \subseteq \mathbb{N}$, let $[X]^n$ denote the size n subsets of X. Suppose that n and m are positive integers and F is a function from $[\mathbb{N}]^n$ to $\{0, \dots, m-1\}$. Then $H \subseteq \mathbb{N}$ is* homogeneous *for F if F is constant on $[H]^n$.*

Theorem 3.7 (Ramsey 1930) *For all positive integers n and m, if F maps $[\mathbb{N}]^n$ to $\{0, \dots, m-1\}$, then there is an infinite set H such that H is homogeneous for F.*

If we fix n and m, we represent the above conclusion as $\mathbb{N} \to [\mathbb{N}]^n_m$.

Theorem 3.7 has a noneffective proof and has been a fruitful example for mathematical logicians. (Jockusch 1972) showed that the noneffective methods in the proof of Theorem 3.7 cannot be eliminated.

Theorem 3.8 (Jockusch)

- *There is a recursive partition of $[\mathbb{N}]^3$ into two pieces such that $0'$ is recursive in any infinite homogeneous set.*
- *There is a recursive partition of $[\mathbb{N}]^2$ into two pieces with no infinite homogeneous set which is recursively approximated.*

Theorem 3.8 gives a good understanding of the definability aspects of Ramsey's theorem, except for the case of partitions of $[\mathbb{N}]^2$. Jockusch posed the following question: Is there a recursive partition of $[\mathbb{N}]^2$ into two pieces such that $0'$ is recursive in any infinite homogeneous set?

More generally, we consider Turing ideals closed under applications of Ramsey's theorem.

Definition 3.9 *A* Ramsey ideal *is a Turing ideal $\mathcal{R}$ such that, if $X \in \mathcal{R}$ and $F : [\mathbb{N}]^2 \to 2$ is recursive in X, then there is an infinite homogeneous set for F in $\mathcal{R}$.*

Seetapun answered Jockusch's question negatively.

Theorem 3.10 (Seetapun). *There is a Ramsey ideal which omits $0'$.*

Thus, Ramsey's theorem is weak with regard to constructing arithmetically defined sets. However, it is strong in a number theoretic sense, as shown by the following theorem of Slaman.

Theorem 3.11 (Slaman). *There is a Π^0_4 statement φ such that*

$$RCA_0 + \mathbb{N} \to [\mathbb{N}]^2_2 \vdash \varphi \quad \textit{and} \quad RCA_0 \nvdash \varphi.$$

(For Theorems 3.10 and 3.11, see (Seetapun and Slaman 1995).)

When we can invoke Ramsey's theorem, the apparatus of infinite sets can be applied to obtain first-order consequences beyond those of RCA_0. It is open whether $RCA_0 + \mathbb{N} \to [\mathbb{N}]^2_2$ proves PA.

Transfinite recursion and Π^1_1 definability Hilbert's tenth problem was to give an algorithm to determine whether a Diophantine equation has an integer solution. One was asked to give a recursive description for a set which came with a natural recursively enumerable presentation. Matijasevič's Theorem states that the recursively enumerable presentation is best possible.

There is an exact second-order analogy to such an undecidability result. One could be given a Π^1_1 set of reals and ask whether it is Borel. In the case of a negative answer, we have a second-order nondefinability result. Examples of this sort are not as well known as those of undecidability and we mention a few of them. An enlightening discussion of these examples and of descriptive set theory in general is given in (Kechris 1995), from which we draw much of our discussion.

Here is a prototypical example, the Cantor–Bendixson theorem. A *perfect* set is a closed set with no isolated points.

Theorem 3.12 (Bendixson 1883) *Suppose that C is a closed subset of the real numbers. Then either C is a countable set or C has a perfect subset.*

In the proof of this theorem one defines an operation on sets, called the Cantor–Bendixson derivative, $A \mapsto A^*$, where A^* is the set of limit points of A. Since the real numbers are separable, $A - A^*$ is countable. Then, one goes by transfinite recursion to define:

$$\begin{aligned} C_0 &= C\,; \\ C_{\alpha+1} &= C^*_\alpha\,; \\ C_\lambda &= \textstyle\bigcap_{\alpha<\lambda} C_\alpha\,. \end{aligned}$$

Again by separability, there is a countable ordinal α such that $C_\alpha = C_{\alpha+1}$. For this α, either C_α is a perfect subset of C, or C_α is empty and so C is countable.

A closed subset of the reals is determined by the set of open intervals with rational endpoints which are contained in its complement and thereby can be regarded as a real number itself. So we can ask, how complicated is the set of uncountable closed sets? By the Cantor–Bendixson theorem, C is uncountable if and only if it has a perfect subset, which is a Σ^1_1 property. Could it be Borel? The answer would be yes if there were a fixed countable α such that for all closed sets C, $C_\alpha = C_{\alpha+1}$. Of course, we chose this example because the answer is no.

The most direct route to verifying the above claim would be to show that the collection of uncountable closed sets is a universal Σ^1_1 set. One would define a recursive function $\{e^*\}$ such that for all e and X, the eth Σ^1_1 statement holds of X if and only if $\{e^*\}(e, X)$ is a code for an uncountable closed set.

There is a less direct route, which seems easier to apply in general. Here ORD is the class of ordinals.

Definition 3.13 *A* Π^1_1-rank *of a set A is a map $\pi : A \rightarrow$ ORD such that the initial segments $A_\alpha = \{x \in A : \pi(x) \leq \alpha\}$ are uniformly Δ^1_1.*

The following is a generalization of a theorem of Spector; see (Kechris 1995, Chapter IV).

Theorem 3.14 *If A is a Π^1_1 set and π is a Π^1_1-rank such that the range of π*

applied to A is unbounded in ω_1, then A is not Borel.

So one can show that the collection of uncountable closed sets is not Borel by showing that there is a Π^1_1 collection of closed countable sets for which the Cantor–Bendixson rank is a Π^1_1-rank and for which that rank is unbounded in ω_1. In fact, the closed countable subsets of the Cantor set have this property.

Now for some other examples. Mazurkiewicz has shown that the set of differentiable functions in $C[0,1]$ is a Π^1_1 set which is not Borel (see (Kechris and Woodin 1986) for a modern proof using ranks); Solovay and Kaufman (independently) have shown that the collection of closed sets of uniqueness is a Π^1_1 set which is not Borel (see (Kaufman 1984) and (Kechris and Louveau 1987)); Slaman and Woodin [unpublished] have shown that the set of countable partial orders which cannot be extended to dense linear orders merely by adding instances of comparability is a Π^1_1 set which is not Borel.

4 The fine hierarchy

Now we turn to the question of the inevitability of our hierarchy of definability.

Definition 4.1 *The* Turing degree *or* degree of unsolvability *of a real A is the set*

$$\{X : A \geq_T X \text{ and } X \geq_T A\}.$$

That is, the equivalence class of A under the relation of equicomputability.

Boolos and Putnam begin their paper (1968) as follows.

> Why the Post–Kleene arithmetical hierarchy of degrees of (recursive) unsolvability was extended into the transfinite is not clear. Perhaps it was thought that if a hierarchy of sufficiently fine structure could be described that would include all sets of integers, some light might be thrown on the Continuum Hypothesis, and its truth or falsity possibly even ascertained.

They then say that the continuum 'has been found to be darker than it was previously known to be' and continue with:

> Nonetheless, the general, 'conceptual' questions of how to extend the extensions of the arithmetical hierarchy so as to include *all* sets of integers and how to assign a degree of unsolvability to every ordinal $<$ (classical) ω_1 in a 'natural' way were interesting in themselves ....

Of course, if there is a $\geq_T$-increasing sequence of reals of length ω_1 such that every real is recursive in some element of the sequence, then the Continuum Hypothesis must hold. But, if one either accepts the Continuum Hypothesis or requires only that every naturally definable real is recursive in some element of the sequence, then there is no reason to deny its existence. In fact, the implicit thesis behind our discussion so far is that there is a sequence of this type which at least covers all of the sets which are defined in commonplace mathematical practice.

Boolos and Putnam proposed a hierarchy of sets, based on Gödel's hierarchy of constructibility. For each α such that there is a real in $L_{\alpha+1} - L_\alpha$, choose an E_α from $L_{\alpha+1}$ so that every real number in $L_{\alpha+1}$ is arithmetically definable from E_α. (They showed that such a choice is always possible.) This hierarchy of sets is flawed in several ways, most of which were acknowledged in the original paper. For example, it only determines sets of level α up to arithmetical equivalence, a failure of detail, and it only encompasses constructible sets, a failure of scope, and it comes unsupported by any compelling evidence of its uniqueness, a failure of predestiny.

As was worked out by Hodes (1980), the missing detail could be provided by replacing the L hierarchy by Jensen's J hierarchy, which provides a finer analysis of constructibility. The sets E_α could be replaced by Jensen's master codes and the sets at each level are naturally determined up to Turing degree, a reasonably fine measure. At the first levels, E_0 would be a recursive set; E_{n+1} would be the universal Σ^0_{n+1} set; if α is isomorphic to a recursive well-ordering of $\mathbb{N}$, then E_α would have the Turing degree of the sets at the αth level in Kleene's hyperarithmetic hierarchy; and if α is ω_1^{CK} the first non-recursive ordinal, then E_α would be $\mathcal{O}$ (Kleene's universal Π^1_1 set).

There is no mandate that restricts a Boolos and Putnam hierarchy to the constructible reals. Every real in L is recursive in $0^\#$, and so $0^\#$ would make a reasonable entry as the ω_1^Lth real in the transfinite hierarchy of Turing degrees. Without an obvious obstruction in sight, the problem of scope is one of implementation.

Even if we could give a detailed hierarchy to satisfy the first two objections, then we would still need to argue that it is the right one, that it is predestined.

Remarkably, there is a conjecture of Martin which, if true, would succeed on all points (see (Steel 1982)). But, before we can formulate it, we need to explain the context in which it is intended.

We have been considering an ω_1 sequence of sets or rather of Turing degrees. We were basing this sequence on the universal sets from the hierarchy of definability embodied by L. Now, we move to a hierarchy of functions from the reals to the reals based on the same hierarchy of relative definability. For example, the Turing degree of the recursive sets is replaced by the identity function, $0'$ is replaced by $X \mapsto X'$, and so forth.

Suppose that A is a set of reals. The game G_A is played between two players, who alternate playing natural numbers $a_1, b_1, a_2, b_2, \ldots$. The first player wins G_A if the resulting infinite sequence belongs to A and otherwise the second player wins. A strategy is a map from finite initial segments of the game to next moves. A strategy is a winning strategy (for one of the players) if it produces a win against every possible play by the opponent. Finally, A is *determined* if there is a winning strategy for one of the two players in G_A. The Axiom of Determinacy (AD) is the statement that every A is determined.

We should note that (Martin 1985) has shown that every Borel game is determined. Martin [unpublished] and (Harrington 1978) have shown that every Π^1_1 game is determined if and only if, for every real X, $X^\#$ exists. Martin, Steel, and Woodin have shown that a sufficiently strong large cardinal hypothesis implies that AD holds in $L[\mathbb{R}]$ (see Martin and Steel 1989). So the assumption of AD is not vacuous.

Definition 4.2

- *A property P holds* almost everywhere *in the Turing degrees if there is a degree d such that, for all x, if $x \geq_T d$, then x satisfies P.*
- *A function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ is* degree invariant *if, for all X and Y, if $X \equiv_T Y$ then $F(X) \equiv_T F(Y)$. Let $\mathcal{I}$ be the collection of degree invariant functions.*
- *For F and G in $\mathcal{I}$, $F \geq_M G$ if $F(X) \geq_T G(X)$ holds almost everywhere in the Turing degree of X.*

We write Martin's conjecture as a conjectured consequence of AD. However, it is more attractively expressed as a conjectured property of all of the functions in $L[\mathbb{R}]$, that is all of the functions which are constructible from the reals, under the assumption that $L[\mathbb{R}]$ is a model of AD.

Martin's conjecture (AD) (i) *If $F \in \mathcal{I}$ and $F \not\geq_M id$, then F is constant almost everywhere.*

(ii) *The order $\geq_M$ is a prewellorder of the set*

$$\{F : F \in \mathcal{I} \text{ and } F \geq_M id\}$$

with the successor of F given by F', where $F' : X \mapsto (F(X))'$.

Each level in the hierarchies of definability which we have discussed is naturally associated with a function from $\mathcal{I}$. Namely, for each real number X and each level of definability there is a universal set relative to X at that level. We have already mentioned X', the universal Σ^0_1 set relative to X, $\mathcal{O}^X$, the universal Π^1_1 set relative to X, and $X^{\#}$, the theory of a set of Silver indiscernibles for $L[X]$. In particular, Martin's conjecture states that any two functions obtained by taking X to such a universal set are comparable. It states in a precise sense that there is only one direction along which to build a hierarchy of definability. Under Martin's conjecture, the Boolos and Putnam and the Hodes hierarchies were initial segments of the only possible one.

Martin's conjecture has another attractive feature: it is essentially proven.

Definition 4.3 *A function F from the reals to the reals is* uniformly degree invariant *if there are functions $t_i : \mathbb{N} \to \mathbb{N}$, for i either 1 or 2, such that whenever X and Y are Turing equivalent with $\{e_1\}(X) = Y$ and $\{e_2\}(Y) = X$, then $F(X)$ and $F(Y)$ are Turing equivalent with $\{t_1(e_1)\}(F(X)) = F(Y)$ and $\{t_2(e_2)\}(F(Y)) = F(X)$.*

In other words, the equivalence of X and Y is transferred to the equivalence of $F(X)$ and $F(Y)$ in a way that does not depend on X and Y. Note that all of the functions that arise as universal sets for notions of relative definability are uniformly degree invariant. Let $\mathcal{I}_u$ be the collection of uniformly degree invariant functions.

Theorem 4.4 (i) (Slaman and Steel 1988) *If $F \in \mathcal{I}_u$ and $F \not\geq_M id$, then F is constant almost everywhere.*

(ii) (Steel 1982) *The set*

$$\{F : F \in \mathcal{I}_u \text{ and } F \geq_M id\}$$

is prewellordered by $\geq_M$ and the successor of F given by F'.

In short, Martin's conjecture is true of the uniformly degree invariant functions.

On the full Martin's conjecture, Slaman and Steel have verified it for all $F \in \mathcal{I}$ such that $id \geq_M F$. They have also shown that the variation of Martin's conjecture to refer to arithmetic invariance rather than Turing invariance is false. In this sense, the failure in detail of the Boolos and Putnam hierarchy also leads to a failure in predestiny.

Even without a proof of Martin's conjecture, in the form that Martin originally proposed it, we regard Steel's Theorem 4.4(ii) as strong evidence that there is only one hierarchy of definability: the one that we have been exploring.

Bibliography

Bendixson, I. (1883). Quelques theorèmes de la théorie des ensembles de points. *Acta Mathematica*, **2**, 415–29.

Boolos, G. and Putnam, H. (1968). Degrees of unsolvability of constructible sets of integers. *Journal of Symbolic Logic*, **33**, 497–513.

Church, A. (1936*a*) A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, **1**, 40–1.

Church, A. (1936*b*) An unsolvable problem of elementary number theory. *American Journal of Mathematics*, **58**, 345–63.

Cohen, P. J. (1966). *Set theory and the continuum hypothesis.* W. A. Benjamin, New York.

Friedman, H. (1975). Some systems of second-order arithmetic and their use. In *Proceedings of the International Congress of Mathematicians*, Vol. 1, pp. 235–42. Canadian Mathematical Congress.

Gaifman, H. (1964). Measurable cardinals and constructible sets. *Notices American Math. Soc.*, **11**, 771.

Gödel, K. (1931).Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatsh. Math. Phys.*, **38**, 173–98. Reprinted as: On formally undecidable propositions of *Principia Mathematica* and related systems I. In *Kurt Gödel: collected works*, Vol. I (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort). Oxford University Press, New York, 1986.

Gödel, K. (1932). Über Vollständigkeit und Widerspruchsfreiheit. *Ergebnisse eines mathematischen Kolloquium*, **2**, 12–13. Reprinted as: On completeness and consistency. In *Kurt Gödel: collected works*, Vol. 1 (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort), pp. 235–7. Oxford University Press, New York, 1986.

Gödel, K. (1938). The consistency of the Axiom of Choice and of the generalized Continuum Hypothesis. *Proc. Nat. Acad. Sci. USA*, **24**, 556–7.

Gödel, K. (1965). On undecidable propositions of formal mathematical systems. In *The undecidable. Basic papers on undecidable propositions, unsolvable problems and computable functions* (ed. M. Davis), pp. 39–71. Raven Press, Hewlitt, New York.

Harrington, L. A. (1978). Analytic determinacy and $0^{\#}$. *Journal of Symbolic Logic*, **20**, 685–93.

Hilbert, D. (1901–1902). Mathematical problems. *Bull. American Math. Soc.* (*N. S.*), **8**, 437–79.

Hodes, H. T. (1980). Jumping through the transfinite: the master code hierarchy of Turing degrees. *Journal of Symbolic Logic*, **45**, 204–20.

Jensen, R. B. and Solovay, R. M. (1970). Some applications of almost disjoint sets. In *Mathematical logic and foundations of set theory* (ed. Y. Bar-Hillel), pp. 84–104. North-Holland, Amsterdam.

Jockusch, Jr., C. G. (1972). Ramsey's theorem and recursion theory. *Journal of Symbolic Logic*, **37**, 268–80.

Jockusch, Jr., C. G. and Soare, R. I. (1972). Π^0_1 classes and degrees of theories. *Trans. American Math. Soc.*, **173**, 33–56.

Kaufman, R. (1984). Fourier transforms and descriptive set theory. *Mathematika*, **31**, 336–9.

Kechris, A. S. (1991). Amenable equivalence relations and the Turing degrees. *Journal of Symbolic Logic*, **56**, 182–94.

Kechris, A. S. (1995). *Classical descriptive set theory.* Springer-Verlag, Heidelberg.

Kechris, A. S. and Louveau, A. (1987). *Descriptive set theory and the structure of sets of uniqueness.* London Mathematical Society Lecture Note Series, Vol. 128. Cambridge University Press.

Kechris, A. S. and Moschovakis, Y. N. (1978). The Victoria Delfino problems. In *Cabal Seminar 76–77* (ed. A. S. Kechris and Y. N. Moschovakis). Lecture Notes in Mathematics, Vol. 689, pp. 279–82. Springer-Verlag, Heidelberg.

Kechris, A. S. and Woodin, W. H. (1986). Ranks of differentiable functions. *Mathematika*, **33**, 252–78.

Kleene, S. C. (1936). General recursive functions of natural numbers. *Mathematische Annalen*, **112**, 727–42.

Kleene, S. C. (1955). Hierarchies of number-theoretic predicates. *Bull. American Math. Soc.* (*N. S.*), **61**, 193–213.

Kleene, S. C. (1987). Kurt Gödel: 1906–1978. *Biographical memoirs*, **56**, 135–78.

Martin, D. A. (1985). A purely inductive proof of Borel determinacy. In *Proceedings of Symposia in Pure Mathematics*, Vol. 42, pp. 303–8. American Mathematical Society.

Martin D. A. and Steel, J. R. (1989). A proof of projective determinacy. *Journal American Math. Soc.*, **2**, 71–125.

Matijasevič, Y. (1970). Enumerable sets are diophantine (in Russian). *Doklady Academy Nauk, SSSR*, **191**, 279–82. Translation in *Soviet Math Doklady*, **11** (1970), 354–7.

Post, E. L. (1936). Finite combinatory processes. Formulation I. *Journal of Symbolic Logic*, **1**, 103–5.

Ramsey, F. P. (1930). On a problem in formal logic. *Proc. London Math. Soc.* (3), **30**, 264–86.

Seetapun, D. and Slaman, T. A. (1995). On the strength of Ramsey's theorem. *Notre Dame J. Formal Logic*, **36**, 570–82.

Shoenfield, J. R. (1961). The problem of predicativity. In *Essays on the foundations of mathematics* (ed. Y. Bar-Hillel, E. J. J. Poznanski, M. O. Rabin, and A. Robinson), pp. 132–9. Magnes Press, Hebrew University, Jerusalem.

Shoenfield, J. R. (1995). The mathematical work of S. C. Kleene. *Bull. Symbolic Logic*, **1**, 9–43.

Slaman, T. A. and Steel, J. R. (1988). Definable functions on degrees. In *Cabal Seminar 81–85* (ed. A. S. Kechris, D. A. Martin, and J. R. Steel). Lecture Notes in Mathematics, Vol. 1333, pp. 37–55. Springer-Verlag, Heidelberg.

Steel, J. R. (1982). A classification of jump operators. *Journal of Symbolic Logic* **47**, 347–58.

Turing, A. M. (1936). On computable numbers with an application to the Entscheidungsproblem. *Proc. London Math. Soc.* (3), **42**, 230–65. A correction, **43**, 544–6.

Department of Mathematics
University of California
Berkeley
CA 94720-3840
USA.
email: slaman@math.berkeley.edu

14

True to the pattern

Gianluigi Oliveri

1 Introduction

The aim of this chapter is to sharpen and defend the view advocated by Wittgenstein in *Philosophical Investigations* that an aspect is not '[...] a property of the object, but an internal relation between it and other objects'.[1]

This conception of aspects is philosophically very important for the following reasons.

First, it provides a view of experience which differs from that inherited by a large part of the post-Kantian tradition.

Secondly, it can be used to establish that aspects (or patterns, as I from here will interchangeably call them) are *real* even though they are neither objects nor properties of objects.

Thirdly, it shows that, if mathematics is a science of patterns, the conception of truth which fits best with it is that of Aristotle/Tarski.

The first section of this chapter will contain a discussion of the traditional views held by Kant, and by much of the post-Kantian tradition, on the exclusively *intellectual* function of concepts, that is, the function of concepts is contributing to understanding by means of judgements. I will express this in Wittgensteinian terminology by saying that concepts, for Kant, are not involved in *seeing*, but in *interpreting*.

In §2, I shall analyse the applicability of the notions of seeing and interpreting to the way in which mathematical patterns are given to us. I will show how unsatisfactory the two above-mentioned notions are to this end and will argue that *aspect seeing* is a characteristic of the imagination that imposes structure over sensory input not by means of Kantian *a priori* intuitions (of space and time), but by means of concepts.

In §3, I shall examine the concept of internal relation, a concept which will enable me to give a more satisfactory characterization of mathematical experience.

In §4, I shall study the impact that the points discussed in the previous sections have on our conception of the metaphysics of experience. In particular, I will attempt to show that the rôle performed by concepts in structuring imagination does not commit one to be an idealist.

Finally, in §5, I will argue in favour of the adequacy of the Aristotle/Tarski theory of truth to mathematical theories.

2 Traditional Kantian teachings concerning concepts and perception

For Kant, the rôle of concepts is that of producing understanding by means of the activity of *combination* that they perform over the manifolds independently given (independently of concepts) through intuitions.[2] (What Kant means by intuition is 'That representation which can be given prior to all thought ...'.)[3]

This feature of Kant's ideas about the sharp separation existing in the way that human reason operates between perception and concepts emerges very strongly from the following passage in which Kant compares human understanding with Divine understanding:[4]

> ... were I to think an understanding which is itself intuitive (as, for example, a divine understanding which should not represent to itself given objects, but through whose representation the objects should themselves be given or produced), the categories would have no meaning whatsoever in respect of such a mode of knowledge. They are merely rules for an understanding whose whole power consists in thought, consists, that is, in the act whereby it brings the synthesis of a manifold, given to it from elsewhere in intuition, to the unity of apperception—a faculty, therefore, which by itself knows nothing whatsoever, but merely combines and arranges the material of knowledge, that is, the intuition, which must be given to it by the object.

If my interpretation is correct, for Kant, concepts do not play any rôle in perception, whose sole factors are the objects belonging to the external world, the senses and the *a priori* pure intuitions of space and time. Once perception takes place, concepts represent the conditions according to which human understanding can be generated.

However, it is extremely important to point out that even though concepts, for Kant, have no part in perception, they are considered by him as *a priori* conditions of experience.[5] There is no contradiction generated here by the opinions expressed above concerning the function of concepts, because, for Kant, there is much more to experience than mere perception.[6] These ideas of Kant remained very influential for quite some time. Many analytical philosophers took on board the Kantian tenet that concepts have nothing to do with perception reformulating it in a more modern terminology. An example of this is given by the traditional distinction operated within analytical philosophy between sentences that are *observational* and sentences that are not.

The factor that is common to the several definitions provided by different philosophers of what an observation sentence is is that 'A sentence S is *observational* just in case S is not theory-laden'. The sentence 'X is a microscope' is theory-laden, because to understand what a microscope is we need to know optics, etc., therefore, observation sentences ought to be sentences that we can understand independently of any theory. In a sense, observation sentences ought to be reports of our observation rather than of our interpretation, reports about what we see and not reports about how we interpret what we see.

Such a Kantian distinction between observation and non-observation sentences has been attacked from many quarters. It has been, in particular, attacked by philosophers of science such as Kuhn who believe that there are no basic facts which are observed/perceived by everybody in the same way, but that the adoption of a particular scientific theory affects also what we perceive:[7]

> Since remote antiquity most people have seen one or another heavy body swinging back and forth on a string or chain until it finally comes to rest. To the Aristotelians, who believed that a heavy body is moved by its own nature from a higher position to a state of natural rest at a lower one, the swinging body was simply falling with difficulty. Constrained by the chain, it could achieve rest at its low point only after a tortuous motion and a considerable time. Galileo, on the other hand, looking at the swinging body, saw a pendulum, a body that almost succeeded in repeating the same motion over and over again ad infinitum. And having seen that much, Galileo observed other properties of the pendulum as well and constructed many of the most significant and original parts of his new dynamics around them. From the properties of the pendulum, for example, Galileo derived his only full and sound arguments for the independence of weight and rate of fall, as well as for the relationship between vertical height and terminal velocity of motions down inclined planes. All these natural phenomena he saw differently from the way they had been seen before.

However, if Kuhn, and others, succeeded in showing that there are no such things as observation sentences, they did not provide us with a new theory of experience which might recognize and give an account of the rôle played by concepts in perception. This latter problem is, in my view, at the heart of the second part of Wittgenstein's *Philosophical Investigations* and will be the object of analysis of what follows in this chapter.

3 Seeing or interpreting?

Budd, in his interesting analysis of Wittgenstein's position on *aspect-seeing*, reaches the conclusion that, for Wittgenstein:[8]

> ... the concept of seeing an aspect *lies between* the concept of seeing colour or shape and the concept of interpreting: it resembles both of these concepts, but in different respects.

The concept of seeing an aspect, says Budd, is similar to that of seeing a colour, because when we say that we see something as a cube or as a wire-box it does not make any sense to ask whether what we are seeing is true or false.

We are not making a conjecture, we are having a perception and reporting it. In other words, there is no being right or wrong involved here.

Moreover, 'seeing an aspect' and 'seeing a colour' are terms that refer to *states*. Both seeing a colour and seeing an aspect have *duration*.

On the other hand, seeing an aspect is also analogous to interpreting, in that seeing an aspect is subject to the will. We can change at will the aspect we perceive of a particular shape simply by concentrating our attention on certain particulars rather than on others as in the well-known duck–rabbit example.

To this, I would add that, for Wittgenstein, a further similarity existing between seeing an aspect and interpreting is represented by the fact that, in order to perceive an aspect, we need to have learned something. We can imagine a baby perceiving a red object, but it does not make sense to say that the baby perceives something as a cube.[9]

According to Budd's interpretation, we ought to say that, for Wittgenstein, human reason performs a third type of activity besides imagining and judging: seeing aspects.

If I understand Budd correctly, seeing an aspect ought to be, for Wittgenstein, a phenomenon that, having characteristics in common both with seeing and interpreting, ought to be described, in the absence of further qualifications, as a kind of *looking + thinking*. However, this conclusion is explicitly rejected by Wittgenstein:[10]

> Is being struck looking plus thinking? No. Many of our concepts *cross* here.

What I take to be one of the clearest accounts given by Wittgenstein of the phenomenon he calls 'dawning of an aspect' is given by the following passage:[11]

> The colour of the visual impression corresponds to the colour of the object (the blotting paper looks pink to me, and is pink)—the shape of the visual impression to the shape of the object (it looks rectangular to me, and is rectangular)—but what I perceive in the dawning of an aspect is not a property of the object, but an internal relation between it and other objects.

From an analysis of the quotation above it seems clear to me that, whatever an internal relation might be, i) it is what we perceive in the dawning of an aspect, ii) what we perceive in the dawning of an aspect is *not* a property of the object.

Some of the consequences I can derive from i) and ii) are that, for Wittgenstein, seeing an aspect is part of the faculty of imagination, i.e. we *perceive* an aspect, and secondly, an aspect not being a property of an object, such a perception is possible, as expressed in note 9, because the person who has it can do, has learnt, is master of certain techniques, i.e. understands certain concepts.

The correctness of this view is particularly visible in the case of mathematical aspects (or patterns)[12], as we see from the following example.

In science we very often come across the problem of solving a system of linear equations, for example the system:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 1 \\ 2x_1 - x_2 + x_3 &= 2 \\ 4x_1 + x_2 - x_3 &= 1\,. \end{aligned}$$

Now the most powerful way of attacking this problem so far found is given by:

i) 'seeing the coefficients of the system as a matrix', i.e.,

$$A = \begin{pmatrix} 1 & -2 & 1 \\ 2 & -1 & 1 \\ 4 & 1 & -1 \end{pmatrix} ;$$

ii) examining the augmented matrix:

$$(A \mid b) = \left(\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ 2 & -1 & 1 & 2 \\ 4 & 1 & -1 & 1 \end{array} \right) ;$$

and

iii) applying concepts and results of matrix theory to find out whether the system of equations has solutions, etc.[13]

If we carefully consider what happens when we are faced with a situation like that described in the example above, we realize that we can *suddenly* see the coefficients of the system of linear equations as a matrix and that such an aspect is not a property of the system of equations, as, instead, would be that individuated by the proposition 'The term on the right-hand side of the equality sign of the first equation (from the top) is 1'. (If we change the system of representation, we will not see the coefficients of the system of linear equations as a matrix, but 1 will always be at the right-hand side of the equality sign of the first equation (from the top) belonging to the system.)[14]

On the other hand, it is just as obvious that concepts are here in play as necessary conditions for the aspect/pattern of matrix to become perspicuous when we examine a system of linear equations.

If my analysis is correct, what is revealed in Wittgenstein's discussion of aspect seeing is not his belief in some kind of phenomenon which interpolates between imagination and intellect showing the existence of another faculty of reason besides the two already mentioned, but the interesting fact that *concepts reach very deep.*

4 Internal relations

In the elucidation of Wittgenstein's ideas on seeing an aspect, I have produced in the previous section a crucial notion that remained unexplained: that of internal relation.

Wittgenstein does not say anything about the nature of internal relations in *Philosophical Investigations* and his silence might be evidence in favour of the opinion that his views on the matter held in the *Tractatus*,[15] the *Notebooks*[16] and the *Notes on Logic*[17] remained, to a large extent, unchanged in his later thought.

What I take these views to be can be summed up in the two following theses:

i) a 2-place relation R is *internal* to two objects a and b just in case it is impossible (Wittgenstein says 'unthinkable') that a and b do not stand in relation R to one another,

ii) internal relations (and properties) are what determines the features, structure of a fact.[18]

One of the most important consequences of characteristics i) and ii), within the system of the *Tractatus*, was that internal relations performed a rôle typical of elements of what Wittgenstein called 'logical form', that is, what he thought propositions and reality must have in common for representations of reality to be possible.[19]

The changes that occurred in the later period were mainly related to the metaphysical account of the *formality* of internal relations. What I mean by this is that, although internal relations were still seen by Wittgenstein in the later period not as properties of objects, but as what was part of our system of representation of objects, he had abandoned his early period account of how such representations are possible.

However, this is not the place to carry out Wittgensteinian exegesis, but to produce clarifications concerning the concept of internal relation.

As on other occasions, Wittgenstein's ideas, despite their suggestiveness, do not produce sharp characterizations of the notions analysed.

I find that, on the contrary, G. E. Moore's way of tackling the problem of internal relations, besides capturing characteristics i) and ii) of Wittgenstein's remarks on internal relations, provides us with a sharp and informative definition.

G. E. Moore, in his seminal paper 'External and Internal Relations',[20] produces an application of the modal context to give a characterization of internal relations able to distinguish them from external relations.[21]

Before I begin a discussion of Moore's definition, I must say that Moore does not deal with relations in general, but with what he calls 'relational properties'.[22] Having made these provisos, which do not alter the relevance or the generality[23] of the discussion, I can give Moore's definition of internal relational property:[24]

> Let Φ be a relational property, and A a term to which it does in fact belong. I propose to define what is meant by saying that Φ is internal to A ... as meaning that from the proposition that a thing has not got Φ, it 'follows' that it is *other* than A.

To express this in modern terminology, we can say that a relational property Φ is internal to a term A just in case A has the property Φ and it is *necessary* that for any x, if x does not have the property Φ, then x is different from A.[25] An example of an internal relational property is the following. Let A and B be triangles in a Euclidean plane α and $\Phi(x) :=$ 'x has the same number of angles as B', a relational property. Then clearly Φ is internal to A.

An example of an external relational property is the following. Let $(\mathbb{Z}, +)$ be the algebraic structure obtained when we define $+$ on $\mathbb{Z}$ (a group), $A := 2$ and $\Phi'(x) :=$ 'x is the inverse of -2'. In this case we have that $\Phi'(A)$ is true, because 2 is the additive inverse of -2, but, if we consider the structure $(\mathbb{Z}, \times)$ then 2, in this case, is *not* the inverse of -2.[26]

From the Moorian definition of the notion of internal relational property (and, therefore, also from the corresponding definition of internal relation), we can show how this satisfies condition i) of Wittgenstein's requirements by saying that if Φ is an internal relational property of an object A then it is *necessary* for A to have such a property or, equivalently, it is *impossible* to have A without $\Phi(A)$ being true (Wittgensteinian condition).

The second Wittgensteinian condition (formality) that must be satisfied by an internal relation can also be seen to obtain for internal relational properties in Moore's sense. In fact, since an internal relational property of an object A is a property without which A could not exist, and since A has also many external relational properties which are important to characterize it as an individual, the set of internal relational properties has as elements the necessary conditions for A to be individuated, and these conditions, if they are not also sufficient, can only individuate the general form that something must have if it has to be an A.

Having so clarified the notion of internal relational property (and, therefore, that of internal relation), we must now turn to the problem of checking the claim made that seeing a mathematical pattern can be explained as an act of perceiving an internal relation holding between objects.

Let us consider the example which I gave in §3. Is it possible, given a system of linear equations, that the set of coefficients of the system in the order in which these appear in the equations, does not form a matrix? The answer to this is 'No', because every linear equation in the system contains a finite number of coefficients, the system contains a finite number of equations and there actually is an effective way of constructing the corresponding matrix of coefficients. Therefore, seeing the coefficients of a system of linear equations as a matrix is an act of perceiving an internal relation existing among the entities which compose the system.

Another example would be seeing something as a circle. The aspect that would become perspicuous to us of an object which looks like a circle is that the set of border points of the object is made of entities which look as if they have the same distance from the centre point of the object. Of course, once again we are perceiving an internal relation existing between different sets of points (centre point and border points).

5 The metaphysics of experience

In §2, I showed that *seeing a pattern* (or an aspect) is what reveals that concepts play a fundamental rôle in perception as well as in the intellectual activity of making judgements.

However, the discussion I conducted there did not shed any light on the function that concepts play in perception, that is, that discussion did not clarify whether we have to consider them as supplementary senses or as what produces magnifications and extensions of the senses, etc.

But, perhaps, some of the investigations contained in §3, where I studied the correctness of the Wittgensteinian assertion that *seeing something as* is a perception of an internal relation existing among objects, will enable me to address successfully this problem.

In the example I gave there of seeing a mathematical pattern (aspect),[27] we can very clearly identify what concepts do in perception. The example presents us with the purest form of organizational function that concepts perform in structuring representation.

If we observe the system of equations, the *seeing* part of the act of seeing the coefficients of the equations belonging to the system as a matrix has no relevance whatsoever, in the sense that we can conceive of this very aspect/pattern dawning on a blind man acquainted with matrix theory and with systems of linear equations.

Now the fact that we can safely generalize this interpretation to the *seeing* of any mathematical pattern leads us to conclude that, although aspect-seeing is a phenomenon characteristic of perception, it does not belong to the sensory part of it. It is what brings purely sensory input into a manifold and the extremely interesting feature of this *phenomenon* (in the Kantian sense) is that concepts perform the structuring/organizational rôle which brings sensory input into a manifold.

According to the above interpretation, concepts, besides presiding over the judgement-making activity of the intellect, perform a function very similar to that of the Kantian *a priori* pure intuitions of space and time. They provide structure (organization) to the sensory input.

The model of experience which derives from these results is very different from that inherited from the post-Kantian tradition. I will list below some of its most remarkable characteristics.

First, although the perceptual and intellectual faculties of reason remain distinct and there is no justification for presupposing the existence of a third faculty of reason interpolating between these two, some of the theoretical vehicles through which such faculties are exercised (concepts) are shared by them.

Secondly, such shared theoretical vehicles are not given *a priori* in the mind, etc., but they are rather the outcome of the cultural activity of human kind. This is a very important point, because it shows that perceptions are influenced by external, non-psychological factors. What this implies is, among other things, that not only language has a very important social aspect, but also so does the mind.

Thirdly, patterns (or aspects) are real, but they are not properties of objects. When I see the coefficients of a system of linear equations as a matrix and then assert 'The matrix which represents the coefficients of the system of linear equations... in the order in which they occur in the system is ... ', the sentence I assert is true or false independently of my possibility of proving it or not. The sentence is true or false according to the structure of the internal relation holding among the coefficients of the equations belonging to the given system. Therefore, the matrix pattern that suddenly dawns on me when I contemplate a system of linear equations is *real.*

On the other hand, the fact that I see the coefficients of a system of linear equations *as a matrix* is not a property of the object, but is a relation and, at that, an internal relation in which its parts stand to one another. If I changed the system of representation then also the aspect that would become perspicuous to me would change, but not the lines and the paper. The object I perceive by my 'bare sight' has all sorts of properties: I draw it with a certain kind of ink on a certain kind of paper, etc.

However, since a discussion of the characteristics and consequences of the new model of experience, which might attempt to do justice to the number and depth of the open problems, deserves much more space than that available in this chapter, I will have to stop here, marvelling once again at the depth of insight present in Wittgenstein's later thought.

6 Truth

If mathematics is a science of patterns, what happens to the notion of mathematical truth? Well, a quick answer to this question is that, whatever patterns might be, since, as I have argued in §5, they are real, the concept of truth which best fits mathematical statements has to be the classical Aristotelian/Tarskian concept of truth.

However, since patterns are admittedly neither objects nor properties of objects, but internal relations which are perceived once a system of representation is in place and objects are given, it is legitimate to ask whether, by accepting such a position, we introduce psychologism into mathematics.

Moreover, if seeing a pattern is a process influenced by learning or by acquired experience, do we also run the risk of introducing full-blown empiricism into mathematics?

Before attempting to answer these two questions, let me say what full-blown empiricism is and why, if psychologism or full-blown empiricism were consequences of the view that mathematics is a science of patterns, this would represent a serious problem for such a position.

What I mean by *full-blown empiricism* in mathematics is the position which states that mathematical concepts are generated from our experience or, to put it in another way, that mathematical truths are inductive truths, i.e., generalizations obtained from particular truths.[28]

This type of empiricism is opposed to what I call *modest* or *quasi-empiricism* which is the view that mathematical theories, and therefore, not single statements, can be *falsified* in the sense that they can be replaced with *better theories*,[29] as in the case of the substitution of ZF set theory for naïve set theory and of the Cumulative Hierarchy of types for the Ramified Theory of Types, etc.

The reason why full-blown empiricism and psychologism are serious problems for any theory they are consequences of is that they are plainly in contrast with essential features of mathematical theories.

Full-blown empiricism is in contrast with the *a priori* nature of mathematical statements (judgements), that is, with the fact that we can provide a justification for asserting that a mathematical statement is true which is independent of experience. To justify that 'if $m, n \in \mathbb{N}$ and $n \neq 0$ then $(m + n') = (m + n)'$, where x' means: the immediate successor of x', it is necessary and sufficient to provide a proof of this statement.

A proof of a mathematical statement has nothing to do with experience in the sense that experience has no bearing on whether or not two mathematical statements A and B are such that B is a logical consequence of A. What has an effect on the existence of a relation of logical consequence between two mathematical statements A and B is the meaning of these two statements for how this is given within the theory to which A and B belong.

Therefore, the meaning of a mathematical statement and the justification of the claim that the statement is true are independent of experience. And if particular experiences, such as those we have when we use a computer to investigate the distribution of certain numbers or to construct models, etc. may serve important heuristic purposes in helping us to bring out patterns clearly or provide displays of particular entities: functions, groups, sets, etc., they have, nevertheless, nothing to do with the meaning and the justification of the truth of a mathematical statement.

If full-blown empiricism falls foul of the *a priori* nature of mathematical statements, the same does not apply to a quasi-empiricist view of mathematics. The reason for this is that for a quasi-empiricist it makes sense to say of a mathematical statement that this is true or false only *within a theory.* This proviso—within a theory—is crucial, because, for a quasi-empiricist, a theory is not a collection of mathematical statements inductively obtained from experience through observation, etc., but, on the contrary, *à la* Popper, the theory, together with its deductive apparatus, comes first.

In other words, the theory is set up before we can even begin to make sense of our perceptions, etc. and carry out our observations. This is what justifies the acceptability of *a priori* true or false mathematical statements within a quasi-empiricist view of mathematical theories and explains the occurrence of the term 'quasi' in the description of such a position. The reason for this is that if the theory is given prior to experience then the relation of logical consequence holding (or not holding) between its statements has obviously nothing to do with experience.

On the other hand, it is the theory as a whole that faces, as Quine put it, the tribunal of experience, in the sense that the theory might always be discarded in favour of a better one. This is what justifies calling this view of mathematical theories 'empirical', because, as Lakatos argued, the possibility of deciding which theory among any two theories is better allows us to write all these theories in the form of a convergent sequence. The fact that the sequence converges shows that there is a reality that these theories are approximating, the experience of which guides us in the construction of new theories.

Psychologism in mathematics is the view that mathematical concepts and relations are founded on mental activity of some description: sensations, memory, mental images and processes, etc.

The reason why psychologism is a very unwelcome consequence for any philosophy of mathematics is that such a position contrasts with the nature of mathematical necessity. Mathematical necessity has nothing to do with mental states. The fact that, 'if $m, n \in \mathbb{N}$ and $m - n > 0$, then $m > n$' has nothing to do with how quickly we grasp that this is the case, or with the duration of the state of understanding, its intensity, etc.

Mathematical necessity has to do with whether or not the truth of a mathematical statement follows from the truth of other mathematical statements.

Obviously, as Frege repeatedly remarked in the *Foundations of Arithmetic*, in judging whether a mathematical statement B follows from a mathematical statement A, only logical considerations are relevant. There might be disputes about what a good logic is or on whether or not something is a proof of a particular statement, but, in all cases, the justification for asserting a particular mathematical statement *is* going to be logical. No experiment or intuition or experience is such as to be able to supplant the function performed by the concept of proof as the only acceptable means for justifying mathematical statements.

If A and B are mathematical statements, then asserting 'B follows from A' is true or false independently of whether the psyche of the utterer works according to Freudian or Jungian or Adlerian psychoanalytic models, that is, independently of the mental process whereby that conclusion is reached.

In particular, proving the statement 'B follows from A' means finding an argument showing that in *any* circumstance in which A is true B is also true. The proviso *any* in 'any circumstance' is evidence of the logical necessity of the relation 'x follows from y', where x and y range over mathematical statements, and of the independence of such relation from considerations which take into account factors other than truth.

Having clarified what full-blown empiricism and psychologism are in the philosophy of mathematics, it is time to see whether the characterization I have provided of patterns has one of those views of mathematics as a consequence.

The fact that a pattern is perceived does not have psychologism as a consequence because seeing something as a triangle does not imply that the concept of triangle has a mental connotation, that is, that it is identifiable with a particular mental process that takes place in our minds whenever we identify or re-identify something as being a triangle.

On the contrary, the necessary condition for identifying something as a triangle or, in other words, to see something as a triangle, is given by our *possession* of the concept *triangle* prior to identification. The crucial factor here is represented by the fact that possessing a certain concept is meant as the ability to perform a number of public activities, such as distinguishing triangles from other geometrical figures or being able to recognize properties of triangles.

The possession conditions of a concept are crucial at this juncture because they clearly point to the fact that having a concept is not a matter of being in a particular mental state, but of being able to do certain things. Therefore, presenting concepts in terms of public possession conditions and showing that the possession of concepts is a necessary condition for pattern seeing dispenses completely with the need to introduce psychological factors into the explanation of seeing mathematical patterns and, in particular, shows that our conception of mathematical activity does not have as a consequence the adoption of psychologism.

But is full-blown empiricism a consequence of my view of patterns? I certainly admit, as I said earlier in this section, that 'seeing a pattern is a process influenced by learning or by acquired experience'. However, the fact that learning and acquired experience influence the phenomenon I called 'pattern seeing' does not imply that mathematical truths are inductive truths, that is, generalizations obtained from particular truths. In fact, we can learn from experience by means of a Popperian process of conjecture–refutation which, in the very first term of the pair of characteristics describing it, bears a strongly anti-inductive mark.

Moreover, the fact that I exclude the possibility of 'seeing something as ... ', in the absence of any previous conceptualization, does not commit me to consider a conceptualization as the framing of a conjecture. (As if, when I perceive something as ..., I make some kind of unconscious hypothesis about what the object is, etc.) But, it speaks in favour of some kind of theoretical commitment to represent the world within a given framework, and this is an attitude which clearly contrasts with the *inductive view* of mathematical truths.

Notes

1. See (Wittgenstein 1983, Part II, p. 212^e).

2. See (Kant 1990, Transcendental Analytic, Section 2 Transcendental Deduction of the Pure Concepts of Understanding, §15, p. 151):

> The manifold of representations can be given in an intuition which is purely sensible, that is, nothing but receptivity; and the form of this intuition can lie *a priori* in our faculty of representation, without being anything more than the mode in which the subject is affected. But the combination (*conjunctio*) of a manifold in general can never come to us through the senses, and cannot, therefore, be already contained in the pure form of sensible intuition. For it is an act of spontaneity of the faculty of representation; and since this faculty, to distinguish it from sensibility, must be entitled understanding, all combination—be we conscious of it or not, be it a combination of the manifold of intuition, empirical or non-empirical, or of various concepts—is an act of the understanding.

3. See (Kant 1990, §16, p. 153).

4. See (Kant 1990, §21, p. 161).

5. See (Kant 1990, §14, p. 126):

> Now there are two conditions under which alone the knowledge of an object is possible, first, *intuition*, through which it is given, though only as appearance; secondly, *concept*, through which an object is thought corresponding to this intuition. It is evident from the above that the first condition, namely, that under which alone objects can be intuited, does actually lie *a priori* in the mind as the formal ground of the objects. All appearances necessarily agree with this formal condition of sensibility, since only through it can they appear, that is, be empirically intuited and given. The question now arises whether *a priori* concepts do not also serve as antecedent conditions under which alone anything can be, if not intuited, yet thought as object in general. In that case all empirical knowledge of objects would necessarily conform to such concepts, because only as thus presupposing them is anything possible as *object of experience.* Now all experience does indeed contain, in addition to the intuitions of the senses through which something is given, a *concept* of an object as being thereby given, that is to say, as appearing. Concepts of objects in general thus underlie all empirical knowledge as its *a priori* conditions. The objective validity of the categories as *a priori* concepts rests, therefore, on the fact that, so far as the form of thought is concerned, through them alone does experience become possible. They relate of necessity and *a priori* to objects of experience, for the reason that only by means of them can any object whatsoever of experience be thought.

6. See (Kant 1990, Book II, Chapter II, 3, Analogies of Experience, pp. 208–9):

> Experience is an empirical knowledge, that is, a knowledge which determines an object through perceptions. It is a synthesis of perceptions, not contained in perception, but itself containing in one consciousness the synthetic unity of the manifold of perceptions. This synthetic unity constitutes the essential in any knowledge of *objects* of the senses, that is, in experience as distinguished from mere intuition or sensation of the senses. In experience, however, perceptions come together only in accidental order, so that no necessity determining their connection is or can be revealed in the perceptions themselves. For apprehension is only a placing together of the manifold of empirical intuition; and we can find in it no representation of any necessity which determines the appearances thus combined to have connected existence in space and time. But since experience is a knowledge of objects through perceptions, the relation [involved] in the existence of the manifold has to be represented in experience, not as it comes to be constructed in time but as it exists objectively in time. Since time, however, cannot itself be perceived, the determination of the

> existence of objects in time can take place only through their relation in time in general, and therefore only through concepts that connect them *a priori*. Since these always carry necessity with them, it follows that experience is only possible through a representation of necessary connection of perceptions.

7. See (Kuhn 1970, Chapter X, pp. 118–9).

8. See (Budd 1987, p. 16).

9. See (Wittgenstein 1983, p. 209^e): 'It is only if someone *can do*, has learnt, is master of, such-and-such, that it makes sense to say he has had *this* experience'.

10. See (Wittgenstein 1983, Part II, p. 211^e).

11. See (Wittgenstein 1983, p. 212^e).

12. See (Oliveri 1997*a*).

13. For details, see (Morris 1986).

14. The conception, held by Wittgenstein, that certain perceptions are influenced by a learning process or, to put it in a slightly different way, by acquired experience is as old as Helmholtz, who wrote:

> ... to many physiologists and psychologists the connection between the sensation and the conception of the object usually appears to be so rigid and obligatory that they are not disposed to admit that, to a considerable extent at least, it depends on acquired experience, that is, on psychic activity. (Richards 1977, p. 238)

and described the very test used by Wittgenstein to discriminate between *seeing* (*innate sensations*, in Helmholtz's terminology) and *seeing something as* (*sensations which are the product of experience*, in Helmholtz's terminology):

> ... nothing in our sense-perceptions can be recognized as sensation [innate] which can be overcome in the perceptual image and converted into its opposite by factors that are demonstrably due to experience. (ibid., p. 240)

The main difference between Wittgenstein's position and Helmholtz's seems to me to lie in the conception of the nature of perceptions which, to use Helmoltz's terminology, are not innate sensations. For Helmholtz, perceptions that are not innate sensations are nothing but unconscious conclusions that we draw on the basis of our interaction with the world. These conclusions, and the concepts we form, have a purely empirical basis. For Wittgenstein, instead, there are no hidden psychological mechanisms which generate concepts and unconsciously derive conclusions from a given set of premisses. Concepts, given in terms of their possession conditions, have a public, social dimension which rests on inventing, learning and mastering a number of conventions and techniques, conventions and techniques which come into being at the same time as the concepts they express.

15. See (Wittgenstein 1981).

16. See (Wittgenstein 1979*a*).

17. See (Wittgenstein 1979*b*).

18. See (Wittgenstein 1981, §4.1221, p. 27):

> An internal property of a fact can also be called a feature of that fact (in the sense in which we speak of facial features, for example).

Also, from (Wittgenstein 1981, §4.123):

> A property is internal if it is unthinkable that its object should not possess it. (This shade of blue and that one stand, *eo ipso*, in the internal relation of lighter to darker. It is unthinkable that *these* two objects should not stand in this relation.) (Here the shifting use of the word 'object' corresponds to the shifting use of the words 'property' and 'relation'.)

19. From (Wittgenstein 1981, §4.1222, p. 26):

> In a certain sense we can talk about formal properties of objects and states of affairs, or, in the case of facts, about structural properties: and in the same sense about formal relations and structural relations. (Instead of 'structural property' I also say 'internal property'; instead of 'structural relation', 'internal relation'. I introduce these expressions in order to indicate the source of the confusion between internal relations and relations proper (external relations), which is very widespread among philosophers.) It is impossible, however, to assert by means of propositions that such internal properties and relations obtain: rather this makes itself manifest in the propositions that represent the relevant states of affairs and are concerned with the relevant objects.

20. See (Baldwin 1993, pp. 79–105).

21. Moore's methodology in tackling this problem anticipates by many years that adopted by Kripke (1980).

22. If $x < y$ is a 2-place relation, which has as extension the set

$$B = \{(a,b) : a, b \in \mathbb{R} \text{ and } a < b\} \quad (B \subseteq \mathbb{R}^2),$$

we call $x < 5$ 'a relational property of the real numbers', because the extension of $x < 5$ is the set $A = \{x : x \in \mathbb{R} \text{ and } x < 5\}$ (in this case $A \subseteq \mathbb{R}$).

23. If you can obtain a relational property from a 2-place relation by substituting an individual constant for a free variable, you can obtain a 2-place relation from a relational property by substituting for an individual constant a variable which differs from the variables (free or bound) occurring in the expression.

24. See (Moore 1922, p. 90).

25. Formally speaking we could put it in this way:

$$(\Phi(A) \land \Box \forall x(\neg\Phi(x) \rightarrow x \neq A)).$$

26. This means that

$$\Diamond \exists x(\neg\Phi'(x) \wedge x = A)$$

is true and, therefore,

$$\Box \forall x(\neg\Phi'(x) \rightarrow x \neq A)$$

is false. Therefore Φ' is an external relation. Moore, in the same paper, had also drawn the conclusion that the propositions formalized by

$$(*) \qquad (\Phi(A) \wedge \Box \forall x(\neg\Phi(x) \rightarrow x \neq A))$$

and

$$(**) \qquad (\Phi(A) \wedge \forall x \Box(\neg\Phi(x) \rightarrow x \neq A))$$

are not equivalent. Indeed, when you consider external relational properties, (**) is true and (*) is false.

27. The mathematical example I gave was that in which, confronted by a system of linear equations, we suddenly see the coefficients of such equations as a matrix.

28. This is the position defended by J. S. Mill and criticized by Frege (1884, §§ 9–10, pp. 12–17).

29. See (Oliveri 1997*b*).

Bibliography

Baldwin, T. (ed.) (1993). *G. E. Moore: selected writings.* Routledge, London.

Budd, M. (1987). Wittgenstein on seeing aspects. *Mind*, **96**, 1–17.

Frege, G. (1884). *Die Grundlagen der Arithmetik.* W. Koebner, Berlin. Translated as *The foundations of arithmetic* (trans. J. L. Austin). Blackwells, Oxford, 1950.

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.

Gödel, K. (1990). In *Kurt Gödel: collected works*, Vol. 2 (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Soloway, and J. van Heijenoort). Oxford University Press, New York.

Kant, I. (1990). *Critique of pure reason* (trans. N. K. Smith). Macmillan, London.

Kripke, S. (1980). *Naming and necessity.* Blackwells, Oxford.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd edn, enlarged). The University of Chicago Press.

Moore, G. E. (1922). External and internal relations. In *G. E. Moore: selected writings* (ed. T. Baldwin), pp. 79–105. Routledge, London.

Morris, A. O. (1986). *Linear algebra.* Van Nostrand Reinhold, Wokingham.

Oliveri, G. (1997*a*). Mathematics. A science of patterns? *Synthèse*, **112**, 379–402.

Oliveri, G. (1997*b*). Criticism and growth of mathematical knowledge. *Philosophia Mathematica* (III), **5**, 228–49.

Quine, W. V. (1993). *Pursuit of truth* (revised edn). Harvard University Press, Cambridge, Massachusetts.

Richards, J. L. (1977). The evolution of empiricism. *British Journal of Philosophy of Science*, **28**, 235–53.

Wittgenstein, L. (1979*a*). *Notebooks 1914–1916* (2nd edn). Blackwells, Oxford.

Wittgenstein, L. (1979*b*). *Notes dictated to G. E. Moore in Norway.* In (Wittgenstein 1979*a*), Appendix II, pp. 108–19.

Wittgenstein, L. (1981). *Tractatus Logico-philosophicus.* Routledge, London.

Wittgenstein, L. (1983). *Philosophical investigations.* Blackwells, Oxford.

Wolfson College
Oxford OX2 6UD
England
email: gianluigi.oliveri@wolfson.ox.ac.uk

PART IV

Sets, undecidability, and the natural numbers

15

Foundations of set theory

W. W. Tait

The topic of our conference is *truth in mathematics*, and certainly the implied question is in part: what constitutes truth in mathematics where, in contrast to the natural sciences, there are no phenomena to be saved? In the fourth century BC, Plato answered this question by observing that we begin with the idea of a certain structure, perhaps derived from experience, and, by analysis (which he called 'dialectic'), we arrive at the principles which define this structure. What is true of the structure then is what can be derived from those first principles. This prescription works very well in the cases of Euclidean geometry, the theory of real numbers, and the theory of finite sets (or of sets of rank less than the least inaccessible cardinal, etc.), where we have ultimately agreed on certain (second-order) axioms which characterize the structures in question to within isomorphism. However, even in these cases, the status of certain propositions about these structures—for example, the Continuum Hypothesis—remains undetermined, pointing to an incompleteness of the laws of the underlying higher-order logic and, in particular, the laws governing the *logical* notion of set (to use the term of (Shapiro 1991)) implicit in the second-order comprehension axiom. In another direction, Gödel's incompleteness theorems yield propositions in predicate logic of any order which are not provable in that system, but which are provable by passing to logic of still higher-order. In this direction we are led to the theory of transfinite ordinals and to set theory; in particular, we are led to the question of the existence of large cardinals. However, in this case, as opposed to the case of Euclidean geometry *et al.*, there does not seem to be just one generally accepted idea of the universe of set theory and of the laws defining it. Thus, perhaps there is no one notion of truth in this subject; there may be different conceptions, each carrying its own notion of truth—as in the case of different geometries.

1 Iterative conception of set theory

But I will speak about a *particular* conception of set theory, which I will call the *iterative conception.* It is explicitly formulated in (Gödel 1947) but is implicit in (Zermelo 1930). In brief, it is the conception according to which sets are the

This is an expanded version of the lecture I prepared for the conference *Truth in Mathematics.* In my lecture, I spoke mainly about the ideas in Sections 1 and 3; but this material will appear in (Tait 1998), and so has been condensed here.

objects in any member of the hierarchy of domains obtained from the null domain by iterating the power set operation $\wp$. This idea presupposes the more primitive, logical notion of set, the one which is formalized in second-order predicate logic, namely the idea of passing from a domain D of objects to the domain $\wp(D)$ of all sets of objects in D. The non-empty domains that can be obtained in this way are, to within isomorphism, the models of the theory T_1 whose language is that of set theory, together with the unary function constant *ran* (for the rank function) added and with the following axioms: *Extensionality*, *Regularity* (in the second-order form asserting that there are no classes which form descending $\in$-chains), *Union*, *second-order Separation*, *Choice*,[1]

$$\forall x[ran(x) = \bigcup\{ran(y) \cup \{ran(y)\} : y \in x\}]$$

(formulated in primitive notation so as to avoid the general assumption of the Axiom of Unordered Pairs) and, with α ranging over the von Neumann ordinals,

$$\forall\alpha[\{x : ran(x) \in \alpha\} \text{ is a set}].$$

For α a von Neumann ordinal in such a domain D, $\{x : \mathrm{ran}(x) \in \alpha\}$ defines the set $R(\alpha)$ of all sets of rank $< \alpha$, which constitutes an initial subdomain of D which we also denote by M_α. Since every domain D is an initial subdomain of another, for example, $\wp(D)$, all domains are of the form M_α.

We may begin simply with the null domain and the operation $D \mapsto \wp(D)$ for obtaining new domains or, what amounts to the same thing, with 0 and the operation $\alpha \mapsto \alpha + 1$ for obtaining new (von Neumann) ordinals. We obtain new such operations by admitting closures under operations that have already been admitted. For example, applied to the operation $\alpha \mapsto \alpha + 1$, we obtain the operation $\alpha \mapsto \alpha + \omega$. The extensions of T_1 that are admitted on the iterative conception are obtained by axioms which express the existence of operations that are obtained in this way. In particular, we shall see that the Axioms of *Infinity* and (*second-order*) *Replacement* follow on this conception. The theory T_1 together with Replacement and Infinity is the full system of impredicative second-order set theory T_2.

The term "iterative conception" has also been used in the literature to refer, not to this *autonomously generated* hierarchy, but to the sets in the hierarchy obtained from the null domain by iterating the operation $\wp$ along *some given* system of ordinals. In this case the question of what sets exist is relative to the system of ordinals with which we start, and there are no grounds for asserting either the Axiom of Infinity or of Replacement on this conception. I will use the term always to refer to the autonomously generated hierarchy.

In (1947), Gödel actually introduced another criterion for accepting new axioms. On p. 265 of the 1964 version of that paper, he writes

> Secondly, however, even disregarding the intrinsic necessity of some new axiom, and even in case it had no intrinsic necessity at all, a probable decision about its truth is possible also in another way, namely, inductively by studying its "success". Success here means fruitfulness

> in consequences, in particular in "verifiable" consequences, i.e., consequences demonstrable without the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover, and make it possible to condense into one proof many different proofs.

This criterion seems to have had some influence among logicians studying models of first-order set theory. But it is difficult to reconcile it with the iterative conception. On the latter conception, the 'intrinsic necessity' of an axiom arises from the fact that it expresses closure under some operation that we have obtained for constructing domains or ordinals. To introduce a new axiom as 'true' on this conception because of its 'success', would have no more justification than introducing in the study of Euclidean space points and lines at infinity because of their success. One may obtain an interesting theory in this way and one worthy of study; but it will not be Euclidean geometry. A 'probable decision' about the truth of a proposition from the point of view of the iterative conception can only be a probable decision about its derivability from that conception. Otherwise, how can we know that a probable decision on the basis of success might not lead us to negate what we otherwise take to be an intrinsically necessary truth?

I want to emphasize that we shall consider *only* the iterative conception of set theory here and, in particular, the question of its strength, measured by what large cardinal axioms can be derived from it. Surely this corresponds to one notion of truth in set theory? I do not want to consider here the question of whether there are other, equally or more satisfactory, notions of truth.

In discussing the iterative conception of set theory further, we shall want to consider formulas of set theory of finite type and their relativizations to a domain.

Definition 1.1

- *The* finite types *are inductively defined by the condition that, if $n \geq 0$ and $\tau_1, \ldots, \tau_n$ are finite types, then $\tau = (\tau_1, \ldots, \tau_n)$ is a finite type.*
- *The* order *of τ is defined to be* 1 *greater than the maximum order of the τ_i.*
- *When $n = 0$, the* objects of type *$\tau = (\,)$ are sets. When $n > 0$, the objects of type τ are relations whose elements have the form $(t_1, \ldots, t_n)$, where t_i is of type τ_i for $i = 1, \ldots, n$. Objects of type $((\ldots(\,)\ldots))$ of order n (i.e., with n pairs of parentheses) will be called* classes of order *n.*
- *The* formulas *are built up by means of the propositional connectives and quantification of variables of arbitrary finite type from atoms of the form $x \in y$, where x and y are of type $(\,)$, or $(X_1, \ldots, X_n) \in Y$, where $n > 0$, Y is of type $(\tau_1, \ldots, \tau_n)$ and X_i is of type τ_i for $i = 1, \ldots, n$. The formulas $(x) \in Y$ and $(X) \in Y$ are written simply as $x \in Y$ and $X \in Y$, respectively.*
- *The* order *of a formula is the maximum order of the bound variables in it (though it may contain free variables of higher order).*

- *When φ is a formula, then its* relativization φ^β *to $R(\beta)$ is the result of restricting the bound variables of type $()$ in φ to the set $R(\beta, ()) = R(\beta)$ of sets of rank $< \beta$ and, for $n > 0$, the bound variables of type $\tau = (\tau_1, \ldots, \tau_n)$ to the set $R(\beta, \tau) = \wp(R(\beta, \tau_1) \times \cdots \times R(\beta, \tau_n))$.*

So when we refer to T_1 or T_2, we are not referring to a second-order theory: its *axioms* are indeed second-order, but the framework of the theory we are considering is not second-order predicate logic, but rather predicate logic of finite order. (So, for example, in the logical comprehension axiom, yielding second-order classes $\{x : \varphi(x)\}$, φ may be of arbitrary finite type.) We could of course consider formulas of transfinite type; but I will avoid that complication here. Generally, in considering set theory of finite order, it is sufficient to introduce just one type of each order n, namely the type $(\ldots()\ldots)$ of classes of order n, since other objects of order n can be coded by classes of order n. But we need to consider the wider class of formulas, because in §4 we shall be interested in a special kind of formulas, the *positive formulas* (see Definition 4.1 below), and the coding in question does not preserve the property of being positive. However, for more immediate purposes, it will be useful to recall the classification of the set of formulas all of whose variables of each order $n > 1$ range over classes of order n. (Compare our definition with Definition 2.1 of (Slaman 1998).)

Definition 1.2 *Let $n \geq 1$.*

- *A formula of order $\leq n$ is called a Π^n_0 formula and a Σ^n_0 formula.*
- *A Π^n_{m+1} formula is one of the form $\forall Y \psi(Y)$, where ψ is a Σ^n_m formula and Y is a class variable of order $n+1$.*
- *A Σ^n_{m+1} formula is the negation of a Π^n_m formula.*

When X is first-order, let $X^\beta = X$. When X is second-order of type $((), \ldots, ())$, X^β denotes $X \cap (R(\beta) \times \cdots \times R(\beta))$. From the point of view of $R(\beta)$, X denotes X^β. For example, when X and Y are second-order classes, then X and Y are equal relative to $R(\beta)$ just in case

$$\forall z \in R(\beta)[z \in X \longleftrightarrow z \in Y].$$

So when $A, \ldots, B$ are of order 2 and those among them which are sets are in $R(\beta)$, then $\varphi^\beta(A^\beta, \ldots, B^\beta)$ expresses the truth of $\varphi(A, \ldots, B)$ in $\langle R(\beta), \in \rangle$ or, as we shall say, in $R(\beta)$. Up to now in the development of the iterative conception of set theory, the engine of iteration has been restricted essentially to the *reflection principle*,

$$\forall X[\varphi(X, \ldots, Y) \longrightarrow \exists \beta \varphi^\beta(X^\beta, \ldots, Y^\beta)], \tag{1.1}$$

where $X, \ldots, Y$ are variables of order at most 2 and are the only free variables in $\varphi(X, \ldots, Y)$. When φ is Π^n_m and k is the maximum order of its free variables, then we denote the principle (1.1) by

$$RF(n, m, k).$$

Definition 1.3

- *An ordinal γ is* φ-indescribable *if* (1.1) *holds in* $R(\gamma)$.
- *If Θ is a class of formulas, then γ is* Θ-indescribable *if it is φ-indescribable for each $\varphi \in \Theta$ containing only free variables of orders ≤ 2.*
- *γ is* totally indescribable *if it is Θ-indescribable, where Θ is the class of all formulas of set theory of finite type.*

So γ is Π^n_m-indescribable if $RF(n,m,2)$ holds in $R(\gamma)$. We shall sometimes, for simplicity, apply these definitions also to the 'totality' Ω of all ordinals, as though it were an ordinal. Thus, the principle $RF(n,m,2)$ is equivalent to the assertion that Ω is Π^n_m-indescribable. However, the reference to Ω is simply a convenience and can always be eliminated.

2 First- and second-order reflection

In this section, we shall discuss the strength of $RF(1,m,1)$ and $RF(1,m,2)$, before considering more precisely how they may be derived from the iterative conception. Gödel seems to have accepted $RF(1,m,k)$ for $m \leq 1$ and $k \leq 2$, at least, as intrinsically necessary on the iterative conception, since he takes, not only the axioms of ZF to follow from that conception, but also the existence of (strongly) inaccessible cardinals and Mahlo cardinals. (See page 264, and, in particular, footnote 20 of (Gödel 1947).) These two kinds of cardinal will be defined below.

Applications of $RF(1,0,1)$ to $\exists y[x = y]$, which we shall abbreviate by $x \in V$, and then to

$$\forall\alpha\exists\beta[\alpha \in \beta \wedge x \in V]$$

imply in T_1, first, that every ordinal has a successor (from which the Axiom of Power Set follows in T_1), and then, that every ordinal is less than some limit ordinal; and so, in particular, they imply the Axiom of Infinity. Let $\varphi(x,Y)$ be the formula

$$Y \text{ is a function with domain } x \text{ and } x \in V.$$

Then $RF(1,0,2)$ applied to $\varphi(x,Y)$ clearly implies the (second-order) Axiom of Replacement. It follows by an application of $RF(1,1,1)$ (with $\varphi(x)$ expressing the conjunction of $x \in V$, Replacement and Power Set) that there is an unbounded class of (strongly) inaccessible cardinals, that is, cardinals κ such that M_κ is a model of T_2. In fact, we obtain something more about this unbounded class of inaccessible cardinals by going to $RF(1,1,2)$.

Definition 2.1

- *A class C of ordinals is* closed *if, for every ordinal β, if $C \cap \beta$ is unbounded in β, then $\beta \in C$.*
- *A class S of ordinals is* stationary *if, for every closed and unbounded class C, $S \cap C \neq \emptyset$.*

Lemma 2.2 *Let $\varphi(X,\ldots,Y)$ be Π^n_m, where $X,\ldots,Y$ are of order ≤ 2. Then $RF(n,m,2)$ implies*

$$\forall X \ldots \forall Y[\varphi(X,\ldots,Y) \longrightarrow \{\beta : \varphi^\beta(X^\beta,\ldots,Y^\beta)\} \text{ is stationary}].$$

For assume $\varphi(A,\ldots,B)$, and let C be a closed unbounded class of ordinals. Apply $RF(n,m,2)$ to

$$[\varphi(A,\ldots,B) \wedge C \text{ is unbounded}]$$

to obtain a β such that $\varphi^\beta(A^\beta,\ldots,B^\beta)$ and $C \cap \beta = C^\beta$ is unbounded in β, so that $\beta \in C$.

So, in particular, it follows using $RF(1,1,2)$ that the class A of inaccessible cardinals is stationary, since the assertion that β is inaccessible is Π^1_1. Applying $RF(1,1,2)$ to the assertion that A is stationary, which is Π^1_1, we obtain a cardinal κ such that $A \cap \kappa$ is stationary in κ, that is, κ is a so-called *Mahlo* cardinal. Using Lemma 2.2, we can iterate this procedure and obtain a stationary class B of Mahlo cardinals, and so cardinals κ such that $B \cap \kappa$ is stationary in κ, that is, *hyper-Mahlo* cardinals; and so on.

$RF(1,1,2)$ yields something more than the existence of Mahlo cardinals, hyper-Mahlo cardinals and the like.

Definition 2.3

- *A* binary tree *is a class* $\mathbf{T}$ *of functions f such that, for some ordinal β, $f : \beta \longrightarrow 2$ and such that, if $f \in \mathbf{T}$ and f has domain β, then f restricted to any ordinal less than β is in* $\mathbf{T}$.
- *A binary tree* $\mathbf{T}$ *is* path-bounded *if, for every function $F : \Omega \longrightarrow 2$, there is an α such that F restricted to α is not in* $\mathbf{T}$.
- $\mathbf{T}$ *is* bounded *if there is an α such that, for all $F : \Omega \longrightarrow 2$, F restricted to α is not in* $\mathbf{T}$.
- *The* binary tree property *is that every path-bounded binary tree is bounded.*

The instance

$$\mathbf{T} \text{ is path-bounded} \longrightarrow \exists\beta\,[\mathbf{T}^\beta \text{ is path-bounded}]^\beta$$

of $RF(1,1,2)$ implies the binary tree property. So, since a cardinal κ is weakly compact just in case it is inaccessible and $R(\kappa)$ has the binary tree property, $RF(1,2,2)$ implies the existence of a stationary class of weakly compact cardinals. We can of course go on to construct a stationary class of *hyper-weakly compact* cardinals; and so on.

The principle $RF(1,0,2)$ not only implies, but is equivalent to, (the conjunction of the axioms of) T_2 in T_1, and $RF(1,1,2)$ not only implies, but is equivalent to, T_2, conjoined with the binary tree property. In other words, inaccessibility is equivalent to Π^1_0-indescribability, and weak compactness is equivalent to Π^1_1-indescribability. Let Y be a second-order class variable. We may restrict the instances of $RF(1,m,2)$ to Π^1_m formulas $\varphi(Y)$ with just Y free, since multiple free variables of maximum order 2 can be coded in T_2 by Y. For $m > 1$, there is a single Π^1_m formula $\varphi(x,Y)$ such that in T_2 every Π^1_m formula containing at most Y as a free variable is provably equivalent to $\varphi(e,Y)$ for some finite ordinal e. It easily follows that, not just for $m = 0$ and 1, but for all m, $RF(1,m,2)$ can be expressed by a single formula ψ_m in T_2. So ψ_m expresses Π^1_m-indescribability in T_2. The formula ψ_m is itself a Π^1_{m+1} formula. So, by Lemma 2.2, $RF(1,m+1,2)$ implies that the class of Π^1_m-indescribable cardinals is stationary. In other words, the class of Π^1_m-indescribable cardinals less than a given Π^1_{m+1}-indescribable cardinal is stationary. So the principle $RF(1,m,2)$ strictly increases in strength as m increases.

A similar argument establishes that, for $n > 1$, the hierarchy of principles $RF(n,m,2)$ is strictly increasing in m. But, for $n > 1$ and $m \geq 1$, it is not so clear that $RF(n,m,2)$ follows from the iterative conception. Anyway, there is in any case a relatively low limit to the cardinals obtained by $RF(n,m,2)$ for arbitrary n and m. Let $[C]^n$ denote the set of all n-element subsets of C, and let $[C]^{<\omega}$ denote the set of all finite subsets of C.

Definition 2.4 *Let $D \subset \kappa$.*

- *$D \longrightarrow (stationary)^n_\lambda$ denotes the following partition property of D: for any function $f : [D]^n \longrightarrow \lambda$, there is a stationary subset S of κ such that $S \subseteq D$ and f is constant on $[S]^n$.*
- *$D \longrightarrow (stationary)^{<\omega}_\lambda$ denotes the following partition property of D: for any function $f : [D]^{<\omega} \longrightarrow \lambda$, there is a stationary subset S of κ such that $S \subseteq D$ and, for each $n < \omega$, f is constant on $[S]^n$.*
- *We may also write $D \longrightarrow (\alpha)^n_\lambda$ or $D \longrightarrow (\alpha)^{<\omega}_\lambda$ meaning that the set S is to be of order type α.*

Notice that the notation suppresses the cardinal κ, which may also be Ω, and which, in each case where it is not explicitly mentioned, will be determined by the context. Now, when κ satisfies even the relatively weak principle

$$\kappa \longrightarrow (stationary)^2_2,$$

then the set of totally indescribable cardinals $< \kappa$ is stationary in κ.

3 Domains and the universe of sets

As an axiom of set theory, (1.1) asserts that, if $\varphi(A)$ is is true in a given model M_κ, then it is true in M_β for some $\beta < \kappa$. But on the iterative conception, we are justified in introducing a given instance of (1.1) as an axiom only by the reflection $(1.1) \longrightarrow \exists\beta(1.1)^\beta$, having first established that (1.1) is true in the universe of all sets—call it M_Ω. But it has been known since (Cantor 1883) that Ω and, consequently, M_Ω cannot be regarded as determinate totalities. Contemporary set theorists frequently write informally as if M_Ω were a model of set theory and, indeed, treat it as if it were a set, except that for some mysterious reason it is not an element of the universe of sets.[2] From their point of view, there is no difficulty with the notion of truth in M_Ω nor with the notion of a higher-order object, say a second-order class A: truth in M_Ω is just truth in a model and A is just a subset of M_Ω which may or may not be (coextensive with) an element of M_Ω. When, as in the case of M_Ω itself, it is not, then it is called a *proper class*. But giving it a name does not really eliminate the mystery of why, when we treat it in all other respects as a set, we nevertheless reject it *as* a set. The paradoxes of set theory on this point of view become the only explanation of why proper classes are not sets—if we admit them as sets, a contradiction arises; but the paradoxes themselves are left with no explanation. I prefer to side with Zermelo (1930) in rejecting the universe of all sets as a well-defined totality and in regarding the paradoxes as arising from the contrary assumption. Indeed, I think that, internal to the iterative conception, there is an explanation of why M_Ω cannot be regarded as a well-defined totality. But, accepting this point of view, the notion of truth in M_Ω requires explanation and we need to explain what we mean by a higher-order object. First, before considering the universe of all sets, we should consider, by way of contrast, the notion of a domain.

3.1 Domains

We have spoken of domains being built up from the null domain by means of iterations of the power set operation $\wp$. But what exactly are we to mean by a 'domain'? It certainly does not coincide with the logical notion of set, since a set in this latter sense is always a set of objects from some prior domain. It is also not useful to identify it with the notion of set that we are analysing in terms of the notion of a domain. In fact, we shall regard domains as a species of *type*. A type is a proposition-like entity (the objects of the type corresponding to the proofs of the proposition): it is given by *introduction rules*, which specify what objects of the given type may be constructed, and by *elimination rules*, which specify how we may reason about them. I shall not go into details about the notion of a type; but some idea of it will be gathered by comparing $M \longrightarrow N$ as an implication between propositions M and N with the interpretation of it as the type of functions from type M to type N. Similarly, compare $\forall x : M.N(x)$ and $\exists x : M.N(x)$ as universal and existential quantifications over the objects of type M with them as types, namely the product, usually written $\prod_{x:M} N(x)$, and the disjoint union or sum, usually written $\sum_{x:M} N(x)$. (Read $x : M$ as 'x is an object of type M'.) In each case, the rules of proof on the logical interpretation are the rules of construction on the ontological interpretation.[3]

The distinction between a domain and a set is now clear: to introduce a type—and in particular, a domain—is to introduce a new kind of mathematical object. The question of whether or not an object is of a given type is always a trivial question (like the question of whether a proof is a proof of a given proposition). The object is *given* as an object of a certain type or else no object has been given at all. In contrast, if s is an object of the domain D and t is a set of objects of D, then the question of whether or not $s \in t$ can be nontrivial.

As a type, the null domain is just the empty type (absurd proposition) and, the domain $\wp(M)$ is just the type $D \longrightarrow TWO$, where TWO is the 2-element type. Corresponding to any *system of ordinals*, that is, a type W with a well-ordering relation defined on it, we may associate with each ordinal $\alpha \in W$ the domain M_α, where M_0 is the null type and $M_{\alpha+1} = \wp(M_\alpha)$. Given ordinals $\alpha \leq \beta$, let $F_{\alpha\beta}$ denote the natural embedding of M_α in M_β. If γ is a limit ordinal in W, then M_γ is the direct limit of the family of maps

$$\langle F_{\alpha\beta} : \alpha \leq \beta < \gamma \rangle .$$

This direct limit operation is not reducible to the ordinary logical operations; but in the Appendix we will establish that it is nevertheless 'proposition-like'.

The problem of constructing domains thus reduces to that of constructing systems of ordinals. Given a system W of ordinals, we may introduce the domain M_W as the direct limit of the domains M_α indexed by ordinals from W. We consider systems W that are introduced in the following way. Let P denote some property of ordinals. If A is a set of ordinals in some system W of ordinals, then $P^*(A)$ is an abbreviation for $\forall\alpha \in A.P(\alpha)$. Relative to P, we may introduce objects of type $W = W_P$ by the introduction rule:

$$A : \wp(W),\ q : P^*(A) \implies S(A,q) : W .$$

Here $S(A,q)$ is the supremum of A. The elimination rule for W is the principle of definition by recursion:

$$p : \forall X : \wp(W) \forall y : P^*(X)[F^*(X) \longrightarrow F(S(X,y))],\ \alpha : W \implies Rp\alpha : F(\alpha)$$

for any property F of ordinals. The meaning of R is given by the recursion equation

$$RpS(A,q) = pAq\,(\lambda\alpha \in A.Rp\alpha).$$

(Here pA denotes $p(A)$, and pAq denotes $p(A)(q)$, etc.) Using definition by recursion, we may define the *transitive closure* $TC(A)$ of a subset A of W as the least subset of W which includes A and such that $S(B,q) \in TC(A)$ implies that $B \subseteq TC(A)$. The system W becomes a system of ordinals when we define the ordering relation by

$$\alpha < S(A,q) \longleftrightarrow \alpha \in TC(A) .$$

Of course, to establish that $<$ is a total ordering, we must define by recursion the notion $\alpha = \beta$ of *equality* between ordinals of type W. Namely, $\alpha \leq \beta$ just in case for every $\alpha' < \alpha$ there is a $\beta' < \beta$ such that $\alpha' = \beta'$; and $\alpha = \beta$ just in case $\alpha \leq \beta$ and $\beta \leq \alpha$.

Of course, not every property P of ordinals can coherently give rise to a system W_P of ordinals. For example, if $P(\alpha) \longleftrightarrow \alpha = \alpha$, then the introduction of W_P would lead to inconsistency. Indeed, consistency requires that $\neg P^*(W_P)$. The iterative conception of set theory yields a plausible restriction on those P for which W_P may be introduced: let $A, \ldots, B$ be of order ≤ 2 and let $\varphi(A, \ldots, B)$ be a true sentence in the universe of sets (assuming, for the moment, that we know what this means). Let $P(\alpha)$ mean that $\varphi(A, \ldots, B)$ is not true in M_α, and let W be the corresponding system of ordinals. It is easy to show that W has a greatest element β and that $\varphi(A, \ldots, B)$ is true in M_β; indeed, β is the least ordinal for which this is true.

3.2 The universe of all sets

The objects of M_Ω are to be all the sets we obtain in any domain, where we abstract from the difference between $x \in M_\alpha$ and $F_{\alpha,\beta}(x) \in M_\beta$ when $\alpha < \beta$. For two sets x and y in M_Ω, $x \in y$ means that, for some β, x has a representative $x' \in M_\beta$, y has a representative $y' \in M_{\beta+1} = \wp(M_\beta)$, and $x' \in y'$. The universe M_Ω then is parasitic off domains. Domains, or, what is the same thing, the corresponding order types β, are what we may construct in developing the iterative conception. The universe M_Ω, on the contrary, can be understood only in terms of what might be constructed on that conception or, better, in terms of the operations we accept or might accept for constructing ordinals or domains; M_Ω can be regarded only as a *potential totality*, partially determined by the operations for constructing ordinals that have been admitted on the iterative conception in any particular argument, but not as a well-defined extension, because there is no characterization of just those operations whose existence is implied by the iterative conception. For a precise characterization of any such set of operations should lead to a new one, namely to ordinals β such that $R(\beta)$ is closed under all of these operations. Moreover, it is uncertain whether there is only one direction in which the iterative conception could develop; perhaps 'at the limit of inquiry', there are divergent paths in its development, yielding different notions of truth.

For this reason, the notion of truth in M_Ω should not be regarded as determined for every sentence and the logic that applies to this universe should be *constructive* logic, not classical. In particular, when we think of the theories T_1 and T_2 and their extensions as theories about M_Ω, then we should take their logical framework to be *intuitionistic* predicate logic of finite type. In the framework of classical predicate logic of finite type, they are not theories of M_Ω; rather, their models are domains in the proper sense.[4]

3.3 Higher-order objects

Since M_Ω cannot be regarded as a well-defined extension, neither can any other second- or higher-order object. Thus, we must regard higher-order objects intensionally, as given by definitions—though, naturally, we do not restrict the definitions to any particular formal system. But when we reflect the object, say the second-order class A, in $R(\beta)$, we do *not* reflect its definition: that is, if $A = \{x : \psi(x)\}$, then A^β is $A \cap R(\beta)$, not $\{x \in R(\beta) : \psi(x)^\beta\}$.

One consequence of this should be noted. Remembering that the logic of M_Ω is constructive, it can happen that the second-order class A, say, is not decidable, in the sense that

$$\forall x(x \in A \vee x \notin A)$$

cannot be derived. For example, the assertion $0 \in A$ may imply the existence of a measurable cardinal, where it turns out that nothing conclusive about the existence of such a cardinal is derivable on the basis of the iterative conception. But A^β for any ordinal β is to be an element of $R(\beta+1)$, and so is to be decidable. In other words, $0 \in A^\beta \vee 0 \notin A^\beta$ should hold; but when $0 < \beta$, this implies that $0 \in A \vee 0 \notin A$. The upshot of this is that, in (1.1), we must restrict the second-order variables, including those among $X, \ldots, Y$, in $\varphi(X, \ldots, Y)$ to decidable objects. A similar remark applies to variables of order < 2, which we shall be considering in the next section. In general an object A of type $(\tau_1, \ldots, \tau_n)$, with $n > 0$, is decidable, just in case,

$$\forall X_1 \cdots X_n[(X_1, \ldots, X_n) \in A \vee (X_1, \ldots, X_n) \notin A]$$

is derivable, where the variable X_i ranges over the decidable objects of type τ_i for $i = 1, \ldots, n$. (First-order objects, that is, sets, are of course decidable.)[5]

However, *having noted this complication in the correct formulation of* (1.1), *which also applies to the reflection principles we shall discuss in the following sections, we shall proceed to ignore it.* The reason that we can afford to do so is that what we are ultimately interested in is what cardinal numbers can be obtained on the basis of the iterative conception. Now, if we have admitted an instance ψ of a reflection principle for M_Ω with the suitable decidability restrictions, then we shall be able to reflect ψ itself to M_κ for some cardinal κ. So any large cardinal property we can obtain by means of the principle will be obtainable for cardinals $< \kappa$. But, relativized to M_κ, the decidability restrictions in ψ become vacuous.

4 Higher-order reflection

As we have noted, relative to $R(\beta)$, the second-order class A is $A^\beta = A \cap R(\beta)$. So the relativization X^β of a third-order class X, for example, is the the class of these. In general, when X is of order > 2 and of type $(\tau_1, \ldots, \tau_n)$, then we should set

$$X^\beta = \{(Y_1^\beta, \ldots, Y_n^\beta) : (Y_1, \ldots, Y_n) \in X\}.$$

With this definition, $\varphi^\beta(A^\beta, \ldots, B^\beta)$, where $A, \ldots, B$ are of arbitrary finite type $\neq (\ )$, expresses that $\varphi(A, \ldots, B)$ is true in $R(\beta)$.

Notice that (1.1) now has meaning for $X, \ldots, Y$ of arbitrary finite type. But there is a problem with the generalized (1.1), even for $X, \ldots, Y$ at most third-order: when U is the class of bounded or of unbounded second-order classes, we have the true sentence $\varphi(U)$ that every class in U is bounded or unbounded, respectively; whereas for every β, $\varphi^\beta(U^\beta)$ is false since U^β is just $R(\beta + 1)$ and, in particular, contains both $R(\beta))$ and the null set. The problem is not merely with generalizing (1.1) to cases in which some of $X, \ldots, Y$ are of order greater than 2, but equally with the case that $\varphi(X, \ldots, Y)$ contains quantifiers of order greater than 2, no matter what the order of $X, \ldots, Y$. For example,

$$\varphi(A) = \exists Y \psi(A, Y)$$

may be true, where A is a second-order class or a set and Y is of order greater than 2, because $\psi(A, B)$ is true for some B. If $\psi^\beta(A^\beta, B^\beta)$ is false for all β, then on what grounds do we infer the existence of a β such that $\varphi^\beta(A^\beta, B^\beta)$? The difference between reflecting formulas containing parameters or bound variables of order at most two, and the reflection of formulas containing parameters or bound variables of higher-order is this: when $\alpha < \beta \leq \Omega$, the second-order structure of M_α is a substructure of M_β; but this is not so for their higher-order structures. In the former case, if φ is a basic (that is, atomic or negated atomic) sentence containing parameters and it is true in M_β, then its relativization to $R(\alpha)$ is true for any α such that the set parameters in φ are of rank less than α. This is clearly so of $s = t$, $s \neq t$, $s \in t$, $s \notin t$, $(s, \dots, t) \in S$, and $(s, \dots, t) \notin S$, when S is second-order. Moreover, it remains true for the higher-order basic sentences $(S, \dots, T) \in R$; but it is not in general true for their negations. Thus, the consideration of (1.1) only for sentences containing no quantifiers of order greater than 2 hides the need for a restriction that is required in the more general case.

Definition 4.1 *A formula is* positive *if it is built up by means of the operations* $\wedge, \vee, \forall$ *and* $\exists$ *from atoms of the form* $x = y$, $x \neq y$, $x \in y$, $x \notin y$, $x \in Y$, $x \notin Y$, $X = Y$ *and* $(X, \dots, Y) \in Z$. *A formula is* negative *if it is built up in the same way, but with '*$(X, \dots, Y) \in Z$*' replaced by '*$(X, \dots, Y) \notin Z$*'.*

So from now on, we shall admit (1.1) in the general case, where both φ and its free variables are of arbitrary finite type and φ is positive. But before discussing this further, we should note and resolve a small difficulty. In the case of a higher-order instance ψ of (1.1), φ is positive, and so ψ itself is negative. Hence, unlike the case of $R(2, m, 2)$, we cannot apply (1.1) to ψ to obtain ordinals β with ψ true in $R(\beta)$. However, there is a negative version of (1.1) which is valid on the iterative conception on precisely the same grounds: For A of order ≤ 2, let $A^{[\beta]} = A^\beta$. For A of type τ and order > 2, define $A^{[\beta]}$ by the condition

$$R(\beta, \tau) - A^{[\beta]} = \{(X^{[\beta]}, \dots, Y^{[\beta]}) : (X, \dots, Y) \notin A\}$$

The operation $A \mapsto A^{[\beta]}$ preserves negative basic sentences, and so on the same grounds that we accept (1.1), we should accept the principle of *negative reflection*

$$\forall X[\varphi(X, \dots, Y) \longrightarrow \exists\beta\varphi^\beta(X^{[\beta]}, \dots, Y^{[\beta]})],$$

where φ is negative and $X, \dots, Y$ are the only free variables in φ. So our difficulty is resolved: if ψ is an instance of positive reflection (1.1), then it is a negative formula without free variables, and so, by negative reflection, we obtain ordinals β such that ψ is true in $R(\beta)$—indeed, we obtain a stationary class of such ordinals.

The information that we have on higher-order applications of positive reflection (1.1) concerns only a special case.

Definition 4.2 *Let* $n \geq 0$.

- *Γ_n will denote the set of all positive formulas of the form*

  $$\forall X_1 \exists Y_1 \cdots \forall X_n \exists Y_n \psi$$

  where ψ is first-order, the X_i are all second-order and, for $i = 1, \ldots, n$, $\exists Y_j$ is a sequence $\exists Z_{j,1} \ldots \exists Z_{j,m_j}$, and the $Z_{j,i}$ are variables of arbitrary types.

- *A subclass B of the cardinal κ is* n-reflective in κ *if*

  $$\forall X \cdots \forall Y[\varphi(X, \ldots, Y) \longrightarrow \exists \beta \in B \varphi^{\beta}(X^{\beta}, \ldots, Y^{\beta})]$$

  for all $\varphi(X, \ldots, Y)$ in Γ_n with $X, \ldots, Y$ of any order ≤ 1. In particular, κ is n-reflective if and only if M_κ satisfies (1.1) *for all such φ.*

In fact, 0-reflection is relatively weak.

Lemma 4.3 *Let κ be a regular uncountable cardinal, let $\varphi(X, \ldots, Y) \in \Gamma_0$ and suppose that $\psi = \varphi(A, \ldots, B)$ is true in $R(\kappa)$. Then there is a closed unbounded subclass C of κ such that ψ is true in $R(\beta)$ for all $\beta \in C$.*

It suffices to take C to be the class of ordinals $\beta < \kappa$ such that $R(\beta)$ is closed under some complete set of Skolem functions for ψ. Just note that, in the case where $(D, \ldots, E) \in F$ occurs in ψ, it occurs positively; and so, if it is false in $R(\kappa)$, it contributes nothing to the truth of ψ in any $R(\beta)$. On the other hand, if it is true in $R(\kappa)$, then it is true in $R(\beta)$ for all $\beta < \kappa$.

The main result that we have concerning higher-order instances of (1.1) is the following:

Theorem 4.4 *For $n \geq 0$, if $D \subseteq \kappa$ is n-reflective, then it has the property that*

$$D \longrightarrow (stationary)_2^{n+1} .$$

In particular, the iterative conception of set theory implies the existence of cardinals κ such that, for every $n < \omega$, $\kappa \longrightarrow (stationary)_2^{<n}$.

The theorem holds even when Γ_n is further restricted to formulas $\varphi(X)$ in which X is of order at most $n + 2$. When $n > 1$, it is unknown whether or not the converse also holds.

There is a very natural property which turns out to be equivalent to n-reflection and which we will use in the proof of Theorem 4.4: a function K defined on $[\kappa]^n$ such that, for $\kappa > \beta_1 > \ldots > \beta_n, K(\beta_1, \cdots \beta_n) \subseteq R(\beta_n)$, is called a *fat n-sequence* on κ. When, more strictly, $K(\beta_1, \ldots \beta_n)$ is always $\subseteq \beta_n$, then K is called a *thin n-sequence* on κ. If K is a thin or fat 1-sequence on κ and $B \subseteq R(\kappa)$, set

$$[K, B] = \{\alpha \in \kappa : K(\alpha) = B \cap R(\alpha)\} .$$

Let K be a fat or thin n-sequence on κ. A subset H of κ is called *homogeneous* for K if there is a $B \subseteq R(\kappa)$ such that $K(\beta_1, \ldots \beta_n) = B \cap \beta_n$ for all

$$(\beta_1, \ldots \beta_n) \in [H]^n \text{ with } \beta_1 > \cdots > \beta_n .$$

If K is a 1-sequence on κ and $B \subseteq R(\kappa)$, then the classes homogeneous for K are precisely those of the form $[K, B]$.

Definition 4.5 *Let $D \subseteq \kappa$.*

- *D is* 0-stationary *if and only if it is stationary.*
- *D is* $n+1$-stationary *if, for every fat 1-sequence K on κ, there is an n-stationary class $\subseteq D$ which is homogeneous for K.*

Theorem 4.6 *Let κ be a regular uncountable cardinal, $D \subseteq \kappa$ and $n < \omega$. Then D is n-reflective if and only if it is n-stationary.*

Notice that being n-stationary and, hence, being n-reflective is a second-order property; so we do not need to invoke negative reflection in order to reflect it—to obtain, for example, a stationary class of n-reflective ordinals.

We sketch the proof of Theorem 4.6. First, a definition. Let τ_0 be the type $(())$ of second-order classes. Set $\tau_{n+1} = (\tau_0, \tau_0, \tau_n)$.

Definition 4.7 *Let $D \subseteq \kappa$.*

- *We define the notion of an n-*box *for D by induction on n. An n-box is of type τ_n.*
    - *A* 0-box *for D is a closed unbounded subset C of κ such that $D \cap C = \emptyset$.*
    - *An* $n+1$-box *for D is an object T of type τ_{n+1} such that for some 1-sequence K on κ, called the* witness for T*:*
        1. *Every element of T is of the form (K, X, S), where $X \subseteq R(\kappa)$, S is an n-box for $[K, X]$.*
        2. *For every $X \subseteq R(\kappa)$, there is an S such that $(K, X, S) \in T$.*
- *Let X be of type τ_0 and T of type τ_n. We define a Γ_n formula $\theta_n(X, T)$ by induction on n.*
    - *$\theta_0(X, T) \longleftrightarrow T$ is an unbounded class of ordinals.*
    - *$\theta_{n+1}(X, T) \longleftrightarrow \forall Y \exists K \exists S[(K, Y, S) \in T \wedge \theta_n([K, Y] \cap X, S)]$.*

Now, to prove that, if D is n-reflective, then it is n-stationary, we show by induction on n that:

- D is not n-stationary if and only if it has an n-box and, if S is an n-box for D, then $\theta_n(D, S)$ is true in $R(\kappa)$;
- If S is an n-box for D, then $\theta_n(D, S)$ is false in every $R(\beta)$ with $\beta \in D$.

In the other direction, assume that D is n-stationary.

- If $n = 0$, then D is n-reflective by Lemma 2.
- Let $n = m+1$, and let $\varphi(A, \ldots, B)$ be a Γ_n sentence which is true in $R(\kappa)$. Then $\varphi(A, \ldots, B)$ is of the form $\forall X \exists Y \psi(X, Y, A, \ldots, B)$, where $\psi \in \Gamma_m$. If $\varphi(A, \ldots, B)$ is false in $R(\beta)$ for all $\beta \in D$, then there is a 1-sequence K

on κ such that $\forall Y \neg\psi(K(\beta), Y, A, \ldots, B)$ is true for each $\beta \in D$. Choose X_0 such that $C = [K, X_0] \cap D$ is m-stationary. There is a Y_0 such that $\psi(X_0, Y_0, A, \ldots, B)$ is true in $R(\kappa)$ and so, by the induction hypothesis, is true in $R(\beta)$ for some $\beta \in C$—a contradiction.

On the face of it, partition properties such as

$$\kappa \longrightarrow (stationary)^n_2$$

do not appear to be susceptible to derivation by reflection (although one should note that both inaccessibility and weak compactness are reflection properties *and*, at the same time, partition properties). An essential step in the proof of Theorem 4.4 is a result of (Baumgartner 1973), which transforms $\kappa \longrightarrow (stationary)^n_2$ into something more like a reflection property.

Definition 4.8 *$D \subseteq \kappa$ is n-*ineffable *if, for every thin n-sequence K on κ, there is a stationary class $\subseteq D$ which is homogeneous for K.*

Before stating Baumgartner's result, it will be of later use to note the following lemma.

Lemma 4.9 *Let K be a fat n-sequence on κ, $n > 0$. Then there is a sequence $\langle G_1, \ldots, G_n\rangle$ of functions such that, for all $X_1, \ldots, X_{n-1} \subseteq R(\kappa)$,*

$$K_i = G_i(X_1, \ldots, X_{i-1})$$

is a fat 1-sequence for $i = 1, \ldots, n$ and, if $\bigcap_{i\leq n}[K_i, X_i]$ is unbounded in κ, then it is homogeneous for K.

The proof is by induction on n. When $n = 1$, there is nothing to prove. Let K be a fat $n+1$-sequence on κ, and define the fat 1-sequence G_1 on κ by

$$G_1(\alpha) = \{\langle \beta_1, \ldots, \beta_n, K(\alpha, \beta_1, \ldots, \beta_n)\rangle : \beta_1 < \alpha\}.$$

For each $X_1 \subseteq R(\kappa)$, define the fat n-sequence $L(X_1)$ on κ by

$$L(X_1)(\beta_1, \ldots, \beta_n) = K(\alpha, \beta_1, \ldots, \beta_n)$$

for all (that is, any) $\alpha \in [G_1, X_1]$ with $\alpha > \beta_1$. If there is no such α, let

$$L(X_1)(\beta_1, \ldots, \beta_n) = \emptyset\,.$$

Now apply the induction hypothesis to $L(X_1)$.

Theorem 4.10 (Baumgartner 1973) *Let κ be a regular uncountable cardinal. Then $D \subseteq \kappa$ is n-ineffable if and only if it satisfies*

$$D \longrightarrow (stationary)^{n+1}_2\,.$$

So, in order to prove Theorem 4.4, it suffices to show that an n-stationary $D \subseteq \kappa$ is n-ineffable for $n > 0$. Again, we sketch the proof.

- For $n = 1$, we need simply to note that every thin 1-sequence on κ is a fat 1-sequence on κ. As a matter for fact, it is easy to show that the converse also holds in this case: 1-ineffability implies 1-stationary. Let C be the class of all inaccessible cardinals and limits of inaccessible cardinals $< \kappa$. Then C is closed and unbounded in κ, since κ is the limit of totally indescribable cardinals. So there is a bijective map $F : \kappa \longrightarrow R(\kappa)$ such that, for each $\lambda \in C$, $F(\lambda) = \lambda$ and the restriction of F to λ is a bijection from λ onto $R(\lambda)$. One easily constructs, corresponding to each fat 1-sequence K on κ, a thin 1-sequence K' such that $K(\alpha) = F[K'(\alpha)]$ for $\alpha \in C$. So $H \subseteq C$ is homogeneous for K' if and only if it is homogeneous for K. It is unknown whether $D \longrightarrow (stationary)^n_2$ implies n-stationary for $n > 1$.
- Let D be $n+1$-stationary, $n > 0$, and K a thin $n+1$-sequence on κ. Let G_1 be as in the proof of Lemma 4.9. There is an $X_1 \subseteq R(\kappa)$ such that $B = [G_1, X_1]$ is n-stationary. Now let $L(X_1)$ be defined as in the proof of Lemma 4.9. Since B is n-stationary, by the induction hypothesis it has a stationary subset A which is homogeneous for $L(X_1)$. So A is clearly homogeneous for K.

So, in particular, Ω is n-ineffable for each $n > 0$ and, therefore, there is a stationary class of cardinals with this property. The properties

$$\kappa \longrightarrow (stationary)^n_2$$

are progressively stronger; an $n+1$-ineffable cardinal has a stationary class of n-ineffable cardinals below it (Baumgartner 1973).

The best bound on the strength of the (1.1) restricted to formulas in Γ_n that is known at the moment has not so far been derived from the iterative conception of set theory.

Theorem 4.11 *Every measurable cardinal is n-reflective for all n.*

Let κ be measurable and let U be a normal ultrafilter on κ. So U contains all closed unbounded subclasses of κ. Theorem 4.11 easily follows from the following lemma.

Lemma 4.12 *Let K be a fat 1-sequence on κ. For every $X \in U$, there is a $Y \in U, Y \subseteq X$, which is homogeneous for K.*

- As we already noted in the sketch of the proof of Theorem 4.10, since κ is the limit of inaccessible cardinals, we may assume that all 1-sequences on κ are thin.
- Let $X \in U$ and let K be a thin 1-sequence on κ. Baumgartner's proof of Theorem 4.10 exhibits a particular $f_K : [X]^2 \longrightarrow 2$ and a closed unbounded class D of κ such that, if f_K is constant on H, then $H \cap D$ is homogeneous for K.

- Rowbottom (1971) proves that, for every $g : [\kappa]^n \longrightarrow 2$, there is an $H \in U$ such that g is constant on $[H]^n$. Take g to be some extension of f_K. Then $H \cap X \cap D \in U$ is homogeneous for K.

5 Appendix: quotient and direct limit types

Let A be a type, and let $g : A \longrightarrow A$, where $g \circ g = g$. The *quotient type* A/g is defined by the introduction rules

$$s : A \Longrightarrow [s] : A/g$$

and the elimination rule

$$t : A/g \Longrightarrow t* : A\,,$$

where $[s]* = g(s)$ and $[t*] = t$. Thus $[s]$ corresponds to the equivalence class of all s' such that $g(s') = g(s)$.

Let $\langle W, < \rangle$ be a well-ordered type and, for each $\alpha : W$, let D_α be a type. For $\alpha \leq \beta$ in W, let $f_{\alpha,\beta} : D_\alpha \longrightarrow D_\beta$ be injective, where $f_{\alpha,\alpha}$ is the identity map on D_α and for $\alpha < \beta < \gamma, f_{\alpha,\gamma} = f_{\beta,\gamma} \circ f_{\alpha,\beta}$. Let $D = \exists\alpha : W \cdot D_\alpha$ and let $g : D' \longrightarrow D'$ be defined by $g(\beta, y) = (\alpha, x)$ for the least $\alpha \leq \beta$ such that y is in the range of $f_{\alpha,\beta}$ and $f_{\alpha,\beta}(x) = y$. We have $g \circ g = g$, and so the quotient type $D = D'/g$ is defined. For $\alpha \in W$, let the maps $f_\alpha : D_\alpha \longrightarrow D$ be defined by $f_\alpha(x) = [(\alpha, x)]$. Clearly D armed with these maps has the required properties of a direct limit of the family of functions $f_{\alpha,\beta}$.

Notes

1. I agree with Zermelo (1930) in regarding the Axiom of Choice in set theory as a consequence of the *logical* principle

$$\forall x \exists y \varphi(x, y) \longrightarrow \exists F[F \text{ is a function} \wedge \forall x \varphi(x, F(x))]\,,$$

which follows from the meaning of the logical constants $\forall$ and $\exists$. For a discussion of this, see (Tait 1994).

2. Of course no *theorem* depends on this assumption. The theorem can always be interpreted in any domain satisfying the axioms being assumed or, since the latter are generally in first-order set theory, in any first-order model of the axioms.

3. This conception of types goes back to (Howard 1980), which has been circulating in manuscript form since 1969.

4. This conception of the universe of all sets is discussed more extensively in (Tait 1997).

5. An alternative approach would be to restrict the higher-order objects in the intuitionistic predicate logic of finite types to decidable relations. In other words, we could adopt a restricted form of the logical comprehension axiom, admitting the relation $\{(X, \ldots, Y) : \varphi(X, \ldots, Y)\}$ only under the condition that it be decidable. On this approach, no restriction on (1.1) is required.

Bibliography

Baumgartner, J. (1973). Ineffability properties of cardinals I. *Colloquia Mathematica Societatus Janós Bolyai*, **10**, 109–30.

Cantor, G. (1883). *Grundlagen einer allgemeinen Mannigfaltigheitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen.* Teubner, Leipzig.

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.

Howard, W. (1980). The formula-as-types notion of construction. In *To H. B. Curry*: *essays on combinatory logic, lambda calculus and formalism* (ed. J. Seldin and J. Hindley), pp. 479–90. Academic Press, New York.

Rowbottom, F. (1971). Some strong axioms of infinity incompatible with the axiom of constructibility. *Annals of Mathematical Logic*, **3**, 1–44.

Shapiro, S. (1991). *Foundations without foundationalism: a case for second-order logic.* Clarendon Press, Oxford.

Slaman, T. A. (1998). Mathematical definability. *This volume*, 233–51.

Tait, W. (1994). The law of excluded middle and the axiom of choice. In *Mathematics and mind* (ed. A. George), pp. 45–70. Oxford University Press.

Tait, W. (1998). Zermelo's conception of set theory and reflection principles. To appear in the proceedings of a conference *Philosophy of Mathematics Today.*

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae*, **16**, 29–47.

Department of Philosophy
The University of Chicago
Chicago, IL 60637
USA
email: wwtx@midway.uchicago.edu

16

Which undecidable mathematical sentences have determinate truth values?

Hartry Field

1 Metaphysical preamble

I will begin by contrasting three metaphysical pictures about mathematics. The first, about which I will have little to say in what follows, is the fictionalist picture. This says that strictly speaking, there are no mathematical entities; still, we can perfectly well reason from premises that postulate such entities, and systematic reasoning from such premises is both intrinsically interesting and highly useful in our practical affairs. On this view, mathematical theories are not literally true. Of course, a fictionalist needs to say something about why these theories are useful if they are not true: in disciplines other than mathematics, the utility of a theory is generally taken as good reason to believe that the theory is at least approximately true, and, if it is not a good reason in the mathematical case, then we need to know what the relevant differences between the mathematical and the non-mathematical are. Myself, I think that there are such relevant differences, and that the fictionalist view can be defended,[1] but I will almost completely ignore it in what follows.

The second picture about mathematics is standard platonism. On the usual version of this view, mathematical theories like number theory and set theory and the theory of real numbers are each about a determinate mathematical domain; or at least, a determinate mathematical *structure*, for there is no need to suppose that isomorphic domains (domains that have the same structure) are distinguishable from a mathematical point of view.[2] Even if a certain sentence in set theory or whatever could not be decided in any mathematical theory that we could have reason to accept, still there is a fact as to whether it is true in the relevant domain or structure. If it is, then it is determinately true, and if not it is determinately false, so it has a determinate truth value despite its being undecidable.

The third picture about mathematics might be called 'plenitudinous platonism'.[3] In one sense of 'platonism' it is an extremely platonist view; in another sense, it is the antithesis of platonism. It is extremely platonist in that it pos-

Thanks to Dan Isaacson and Tony Martin for some helpful questions at the conference, which affected the final version of this chapter.

tulates lots of mathematical objects: at least as many as standard platonism does, and in some loose but intuitive sense, many more. But Kreisel famously remarked that the interesting issue is not mathematical objects but mathematical objectivity, and on the objectivity issue, plenitudinous platonism is virtually indistinguishable from fictionalism.

So what exactly is 'plenitudinous platonism'? Well, roughly what it is, is the view that whenever you have a consistent theory of pure mathematics (that is, a consistent theory that neither postulates non-mathematical objects nor employs non-mathematical vocabulary), then there are mathematical objects that satisfy that theory *under a perfectly standard satisfaction relation.*[4] It is of course simply a theorem of set theory that, whenever you have a consistent (first-order) theory (pure mathematical or not), then there are mathematical objects that satisfy it under some satisfaction relation or other: that is the completeness theorem. What plenitudinous platonism adds is that as long as the consistent theory is purely mathematical, then you do not need to cook up some artificial model out of sets to satisfy the theory; rather, the theory is trivially satisfied by entities that the theory is about and that exist purely by virtue of the consistency of the theory. Indeed, the plenitudinous platonist will add, the idea that you need to cook up some model out of sets to satisfy the theory is only motivated if one thinks that there is a uniquely privileged notion of set. But there are many internally consistent set theories that conflict with each other (differing, for instance, over the size of the continuum); for each one, there are mathematical entities satisfying it; and there is no point in supposing that there is a privileged notion of set such that the entities satisfying each of these theories are all constructed out of the entities satisfying the privileged theory.

I do not pretend that this is a claim of extreme precision, but I do think it has a fairly clear intuitive content. Another imprecise way to put it is as the view that all the consistent concepts of set and membership are instantiated side by side. 'Instantiated side by side' is intended to convey the idea just stressed: we refuse to single out one instantiation as privileged and to regard all others as merely 'unintended models' generated by the completeness theorem. But it also suggests that we could regard all quantifiers over mathematical entities in a mathematical theory as implicitly restricted by a predicate to which all other predicates of mathematical entities in the theory are subordinate. In different mathematical theories the overarching predicate is different; so mathematical theories that appear to conflict with each other when written without their overarching predicates do not really conflict. (It is not necessary to say that there is anything to preclude *meaningfully* quantifying over all mathematical entities at once, without an overarching predicate; one can say just that there is not anything interesting and true to say about so plenitudinous a realm.)

The phrase 'instantiated side by side' suggests that nothing is included under more than one overarching predicate. Actually though it is better to say that there is no mathematical interest to the question of whether things falling under one overarching predicate also fall under another, and the matter can be conventionally decided either way. If sets$_{23}$ are entities that satisfy standard set theory plus the claim that the size of the continuum is $\aleph_{23}$, and sets$_{817}$ are entities that satisfy standard set theory plus the claim that the size of the continuum is $\aleph_{817}$, then it is mathematically uninteresting whether the sets$_{23}$ are included among the sets$_{817}$, or the sets$_{817}$ are included among the sets$_{23}$, or neither inclusion holds. (If an inclusion does hold, or there is overlap, the membership relations $\in_{23}$ and $\in_{817}$ need not coincide on the common domain.) If we want to decide the matter by convention, the best convention is probably to say that neither inclusion holds, simply in order to emphasize that for mathematical purposes neither sets$_{23}$ nor sets$_{817}$ have privileged status.

There is a strong contrast then between plenitudinous platonism and Quinean platonism. Quinean platonism takes as basic some one conception of set, and constructs out of sets so conceived all other mathematical objects: natural numbers, real numbers, and, if we wish, sets corresponding to other conceptions. Most platonists think that Quinean insistence on viewing natural numbers and real numbers as 'really sets' is perverse and arbitrary: why not regard them as perfectly good objects existing on their own? Plenitudinous platonism just goes this anti-Quineanism one better: we should also regard sets that satisfy conceptions other than our own as perfectly good entities in their own right, in no way requiring an explanation in terms of 'our' sets.

Ontologically speaking, then, plenitudinous platonism is highly platonistic, indeed more platonistic than standard platonism: roughly, it postulates multiple mathematical universes where standard platonism (especially Quinean platonism) postulates only one. But methodologically speaking, plenitudinous platonism is quite anti-platonistic (or, as I prefer to say, anti-objectivistic). To illustrate: the usual platonist view is that even after we know that the continuum hypothesis is undecidable from the standard axioms, there is still a serious question as to whether it is true, and we can still find indirect evidence for its truth. The plenitudinous platonist view is that there is no such question: set theory with the continuum hypothesis and set theory with various alternatives to it are all consistent, so all are true of their appropriate domains; and the 'indirect evidence' is simply a matter of exploring the logical implications of each set theory and making aesthetic judgements about their attractiveness and practical judgements about their utility based on these implications. Obviously this is methodologically identical to the fictionalist view, that takes each such set theory to be a fiction and evaluates the fictions on aesthetic and practical grounds.

Actually in saying that plenitudinous platonism has this 'anti-objectivist' methodological consequence, I am being a bit quick. After all, one might adopt the ontological position that there are multiple 'universes of sets' and hold that nevertheless we have somehow mentally singled out one such universe of sets, even though anything we say that is true of it will be true of many others as well. But since it is totally obscure how we could have mentally singled out one such universe, I take it that this is not an option any plenitudinous platonist would want to pursue.

I have tried to spell out plenitudinous platonism in a way that makes it look attractive, but it really does not matter for the rest of the chapter whether you think it even makes sense.[5] For my real interest is not plenitudinous platonism *per se*, but the associated issue about objectivity just distinguished from it. And this issue can arise even in the usual platonist picture of a single universe of sets. That is, if we now take 'standard platonism' to mean simply the usual 'single universe' picture, then the advocate of standard platonism could find himself wondering how we have managed to single out the full universe as opposed to a subuniverse, and the standard membership relation as opposed to some non-standard one. This is just the question that Hilary Putnam raised in (Putnam 1980), from a 'single universe' perspective. Putnam argued in fact that there is no way that we can have managed to determinately single out the full universe of sets and the membership relation on it, and that the incomplete content we have succeeded in giving to 'set' and 'member of' is not enough to determine the truth value of all set-theoretic sentences. In this way, he drew the same anti-objectivistic methodological consequence that the plenitudinous platonist arrives at more directly. In other words, though the usual version of standard platonism has built into it that we do have the set-theoretic universe and the set-theoretic membership relation determinately in mind, this is really not part of standard platonism *per se*; and, if Putnam is right, this usual version of standard platonism cannot be maintained.

I am inclined to think that Putnam is right: that the 'anti-objectivist' methodology is in the end the right consequence for anyone to draw, whatever their ontological views. Perhaps the main advantage of plenitudinous platonism over standard platonism is that (like fictionalism) it leaves little room for disguising this.

2 The objectivity issue

The issue that I will be concerned with, then, is the objectivity issue. If we put the fictionalist option aside, we can formulate the issue as: which of our mathematical sentences have determinate truth values? I will assume that the mathematical sentences we accept are all determinately true (or rather, that those we would continue to accept when any logical errors are filtered out are all determinately true). This seems reasonable, given that the content of our mathematical sentences is determined in large part by which ones we accept, and that there is no reason (independent of fictionalism anyway) to think that such independent determinants of content as there may be exert any pressure toward taking accepted mathematics as untrue. (Roughly speaking, then, accepted mathematics is 'true by convention', or true by the logical consequences of our conventions.[6,7]) Given this, we can rephrase the question as:

(DTV) Which undecidable mathematical sentences[8]
have determinate truth values?

By 'undecidable mathematical sentences' I mean 'mathematical sentences such that neither they nor their negations follow in first-order logic from our fullest mathematical theory'. Platonists in Kreisel's sense—objectivists, I am calling them—assume the answer to be 'all' (all the ones well formulated in our currently best mathematical language, that is); the extreme anti-objectivist position is that the answer is 'none'. Although the remarks at the end of Section 1 might suggest that I favor extreme anti-objectivism, I actually tend to favor an anti-objectivism not quite so extreme; for we will see that our conventions can sometimes make determinately true certain sentences of pure mathematics that are not logical consequences of those conventions.[9]

Two points of clarification are required. First, when I say that certain mathematical sentences might lack determinate truth value, I do not intend to suggest that we must abandon classical reasoning in connection with those sentences. In my view a great many concepts involve some sort of indeterminacy—for instance, vagueness—and as a result many sentences containing them lack determinate truth value. It would cripple our ability to reason if we were prevented from using classical logic whenever indeterminacy might arise. Fortunately, it is not necessary to do so: we can perfectly well say that everyone is either bald or not bald, as long as we add that not everyone is either determinately bald or determinately not bald. What is crucial to the logic of vagueness is not that we give up classical logic but that we add to it a new 'determinately' operator—in effect, a notion of a sentence being determinately true. The same holds in the case of other sorts of indeterminacy. Consequently, standard mathematical reasoning can go unchanged when indeterminacy in mathematics is recognized: all that is changed is philosophical commentaries on mathematics, commentaries such as 'Either the continuum hypothesis is determinately true, or its negation is determinately true'.

Second clarification: it might be protested that it is unclear what should count as 'our fullest mathematical theory', and that this makes 'undecidable' in (DTV) unclear. I agree, but I do not think it matters much for question (DTV): for on any reasonable construal of what counts as our fullest mathematical theory, there will be sentences undecidable in it,[10] and the question is which (if any) such sentences have determinate truth value. We should construe the extreme anti-objectivist as asserting that, if a sentence is undecidable on every reasonable candidate for our fullest theory (and some mathematical sentences surely will meet this condition), then it definitely has no determinate truth value; whereas if it is undecidable on some reasonable candidates but decidable on others then there may be no clear fact of the matter as to whether it has a determinate truth value.

I think that anti-objectivism has considerable plausibility for the typical undecidable sentences of set theory. It has much less plausibility for the undecidable sentences of elementary number theory: these strike almost everyone pre-theoretically as having determinate truth value, though we may not know what it is. I suspect that this feeling arises from the feeling that we have a determinate notion of finitude (that is, of 'finite set', or of the quantifier 'only finitely many', which I abbreviate '$\mathcal{F}$'). Let $\mathbb{N}^+$ be elementary number theory plus 'Every natural number has only finitely many predecessors'. Our fullest mathematical theory includes $\mathbb{N}^+$; and if 'finite' is determinate, all models of this that are not definitely 'unintended' are isomorphic, so all give the same truth value to any number-theoretic sentence A, so A determinately has this truth value by virtue of our commitment to $\mathbb{N}^+$. The incompleteness of formal arithmetic results from the incompleteness of the theory of $\mathcal{F}$; but if the latter is held determinate despite the undecidability of certain sentences in it, the same will hold derivatively of the concept of natural number.

If we do assume the determinacy of finitude, then we can give determinate sense to a conception of semantic consequence for sentences which builds in that '$\mathcal{F}$' gets 'the correct' truth conditions. Then if we call a sentence *f-decidable* if and only if either it or its negation is a semantic consequence (in this sense) of our fullest mathematical theory, any f-decidable sentence will get a determinate truth value. So the key question is: can the assumption that '$\mathcal{F}$' is perfectly determinate be maintained?

There is certainly reason to *hope* that it can: the notion of finiteness is after all a central ingredient in many key notions, such as that of a sentence in a given language (sentences being finite strings of symbols meeting certain conditions) and that of a proof in a given system (proofs being certain finite strings of sentences). Any indeterminacy in the notion of finiteness would doubtless infect the notion of sentence and proof, of logical consequence and logical consistency,[11] and perhaps indirectly even the first-order logical constants themselves (since our understanding of them depends on proof or consequence). Whether such a conclusion would be completely devastating is a question I will return to. But it is clear that we would like to be able to maintain the determinacy of 'finite', and thus of 'natural number'. And I think we can: I will argue that there is a natural account of how our practices might give determinate content to undecidable but f-decidable sentences, an account that does not extend to the typical undecidable sentences of set theory.

3 Putnam's 'Models and Reality' and the concepts of finiteness and natural number

In the first half of (Putnam 1980), Hilary Putnam gave what I think is a compelling argument against the objectivist position in set theory: he argued that, even accepting the component of standard platonism that says there is a single set-theoretic universe $\langle V, E\rangle$, there is nothing in our inferential practice that could determine the truth-value of typical undecidable sentences. (The argument easily extends to an argument against the objectivity of the semantic consequence relation in second-order logic, even if that is interpreted as in (Boolos 1984).[12]) In basic outline, the argument is this:

(Ia) There is nothing in our inferential practice that could determine that our term 'set' singles out the entire set-theoretic universe V rather than a suitably closed subpart of V.

(Ib) Even on the assumption that it singles out the whole set-theoretic universe V, there is nothing in our inferential practice that could determine that our term '$\in$' singles out the membership relation E on V as opposed to some other relation on V that obeys the axioms we have laid down.

(II) The indeterminacy in (Ia) and (Ib) is sufficient to leave indeterminate the truth value of typical undecidable sentences of set theory. (Actually (Ib) by itself would suffice. In many cases (Ia) by itself would also suffice.)

My interest in this part of the chapter is in the scope of this argument: assuming it basically correct, just what undecidable statements does it cover? What are the limits of the semantic facts that our inferential practices might determine? I will try to show how our inferential practices might determine the semantics of 'set' and '$\in$' well enough to make determinate the quantifier 'only finitely many' (when defined in terms of 'set' and '$\in$' in one of the standard ways), and hence determine the truth value of f-decidable but undecidable sentences, including the undecidable sentences of number theory.

Some readers of Putnam have thought that his argument for indeterminacy is based simply on the existence of non-standard models.[13] But if that were so, it would apply as much to non-mathematical language as to mathematical. It seems pretty clear that that is an unattractive consequence: the natural picture is that our practice of accepting and rejecting sentences containing predicates like 'red' and 'horse' and 'longer than' largely determines which of the objects and pairs of objects that we quantify over satisfy these predicates. This seems plausible because our inferential practice with these words includes not only general theoretical principles, but an observational practice which causally ties their extensions down. Other physical predicates, say 'neutrino', are less tied to observational practice, but the theoretical principles governing them include words that are more tied to observational practice, and this does a lot to fix their extension. The *prima facie* problem in the mathematical case is that the theoretical principles of pure mathematics do not tie down predicates like '$\in$' even in this indirect way.

This might suggest that *in the case of mathematical predicates*, indeterminacy of extension follows simply from the existence of non-standard models. If this were so, there would be no hope of exempting the finiteness quantifier from indeterminacy. Let a *grossly non-standard model* of set theory be one in which there are objects y that satisfy 'finite set' even though for infinitely many objects x, the pair $\langle x, y\rangle$ satisfies '$\in$'. (This is equivalent to a non-ω model.) If set theory has models at all, it has grossly non-standard ones, and in these $\mathcal{F}$ gets a non-standard interpretation.

But our inferential practice with terms like 'set' and '$\in$' extends beyond pure mathematics: we use these notions in physical applications. That is, anyone who has learned the notions of set and membership will apply them (and related notions like that of a function, that are usually defined in terms of them) to the physical world. These physical applications of 'set' and '$\in$', not just the mathematical applications, are available to help determine their extensions (given that the extensions of the physical predicates are reasonably determinate). And perhaps these physical applications make determinate the extension of $\mathcal{F}$.

I will argue that if our physical theory includes certain reasonable but not totally obvious 'cosmological assumptions', and if these assumptions are correct, then the extension of $\mathcal{F}$ will indeed be determinate, so that f-decidable sentences will get determinate truth values.

The cosmological assumptions I will use are these:

(A) time is infinite in extent;
(B) time is Archimedean.

More precisely, define $\Phi(Z)$ to mean

> Z is a set of events which (i) has an earliest member and a latest member; and (ii) is such that any two of its members occur at least one second apart.

Then assumption (A) is that there is no finite bound on the size of sets that satisfy Φ; assumption (B) is that *only* finite sets satisfy it.

In accordance with the discussion above, I am going to assume that our physical vocabulary is quite determinate. Indeed, for simplicity, I will assume that it is *completely* determinate, at least relative to a choice of domain of quantification. (I will not need to assume that the domain of our quantifiers is determinate, even when the quantifiers are restricted to physical entities.) More precisely, I will take an interpretation or model of our own language to be 'unallowable' or 'objectively unintended' unless the extension of 'cow' in the model includes precisely those members of the domain of the model that are cows; and similarly for every other physical predicate of the language. As I have said, this is doubtless a little more determinacy than we actually have, but I do not think the over-simplification is harmful in the present context. Reasons for not assuming the analogous determinacy in the mathematical case were given earlier in this section.

Now, let S consist of the above cosmological assumptions plus set theory. (*Impure* set theory, that is, set theory that postulates sets whose members are non-sets, including a set of all non-sets, and including sets defined using physical vocabulary.) From S we can infer the following:

$$(*)\ \mathcal{F}xA(x) \equiv \exists Y\, \exists Z\, \exists f\ [\Phi(Z)\ \&\ Y \text{ contains precisely the } x \text{ such that } A(x)\ \&\ f \text{ is a function that maps } Y \text{ injectively into } Z].$$

I claim that if we accept S, then this consequence of it allows us to extend the determinacy in the physical vocabulary to the notion of finiteness.

I do not deny of course that the theory S has non-standard models, in which certain infinite sets satisfy $\Phi(Z)$ and hence in which certain infinite sets satisfy the predicate 'finite'. However, if assumption (B) is true, any such model must assign a non-standard extension to the formula $\Phi(Z)$; and in particular, it must either contain things that satisfy 'event' which are not events, or it must contain pairs of events which satisfy 'earlier than' or 'at least one second apart' even though the first is not earlier than the second or the two are not one second apart. Either way, the model will violate the constraint on the interpretation of the physical vocabulary: that the extension of such a predicate in the model can only contain things that actually have the corresponding property. In sum: if assumption (B) holds, then *no model of S in which 'event' and 'earlier than' and 'at least one second apart' satisfy the constraints on the interpretation of the physical vocabulary can be one where any infinite sets satisfy $\Phi(Z)$.*[14] But it is easy to see that no pair of an infinite set and a finite set can satisfy

$$\exists f(f \text{ is a function that maps } Y \text{ injectively into } Z)$$

in any model; so no allowable model of S can be one where any infinite sets satisfy 'finite'. And of course there are no models of S where any finite sets fail to satisfy 'finite', since S includes set theory and genuinely finite sets satisfy 'finite' in every model of set theory; so (if assumption (B) holds) *in every allowable model of S, a set satisfies 'finite' if and only if it is genuinely finite.* As a corollary, every number-theoretic sentence gets the same truth-value in every allowable model of S.

This would be cold comfort if assumption (A) were not also true: if it were false, no model meeting the constraints on the interpretation of our physical vocabulary could satisfy set theory plus (*) (since finite sets bigger than any sets that actually satisfy Φ would be constrained to be both finite and infinite); in that case, the interpretation of 'finite' could not be fixed by (*): (B) without (A) is not enough.[15]

The key to the argument, of course, is the assumption that the physical world provides an example of a physical ω-sequence that can be determinately singled out. The cosmological assumptions (A) and (B) are really unduly restrictive, since they entail a specific way in which a physical ω-sequence might be determinately singleable out. Other ways are possible too: for instance, it might be that while time is non-Archimedean, a certain kind of matter exists only in the initial ω-sequence of it (but exists arbitrarily late through that ω-sequence); or it might be space rather than time that provides the physical ω-sequence. However the physical ω-sequence is determinately singled out, it is easy to then use the bridge between the physical and the mathematical that is given in standard (impure) mathematics to make the notion of finiteness fully determinate even in its mathematical applications.[16]

It might be thought objectionable to use physical hypotheses to secure the determinacy of mathematical concepts like finiteness. I sympathize—I just do not know any other way to secure their determinacy. It might be thought *especially* objectionable to use physical hypotheses to secure the determinacy of mathematical concepts like finiteness when those hypotheses are themselves expressed in terms of those mathematical concepts. But recall that in my view, we need not first secure the determinacy of a concept before we use it in reasoning: if that were required, reasoning could never get started. Rather, we can reason classically with our concepts from the start. Indeed, I have claimed that such classical reasoning is not to be called into question should we later discover by such reasoning that the concepts employed in it lack determinacy; it certainly is not to be called into question when (as envisioned here) we discover by such reasoning that the concepts are determinate.

4 Extreme anti-objectivism

I think it is pretty clear that the sort of considerations just given in the case of the theory of natural numbers do not extend to typical undecidable statements of set theory, such as the continuum hypothesis.[17] This does not seem to me in the least disturbing: I do not see any pre-theoretic reason why such statements should be assumed to have determinate truth value. Their lack of determinate truth value seems to me to be fully compatible with accepted methodology in mathematics. I have already pointed out that the recognition of indeterminacy in no way forces us to give up classical reasoning. Also, we can still advance aesthetic criteria for preferring certain values of the continuum over others; we must now view these not as *evidence that* the continuum has a certain value, but rather as *reason for refining our concepts so as to give* the continuum that value, but I do not see this as in violation of any uncontroversial methodological demand.

What might be more disturbing about the considerations advanced here is that the case for the determinacy of finiteness was based on certain 'cosmological hypotheses'. Not, to be sure, on the specific cosmological hypotheses (A) and (B): as noted, these could fail and other cosmological hypotheses be used in their place. But if neither (A) and (B) together, nor any suitable surrogate for them, were correct, then the argument that undecidable but f-decidable statements have determinate truth value would break down. And this may seem disturbing: most of us feel that we have a perfectly definite conception of finiteness, which gives a definite truth value to f-decidable statements even if we do not know what it is. The consequences of giving this up seem quite radical: I observed earlier that if we do not have a determinate conception of finitude, then we do not have a determinate conception of *formula of a given language*, or *theorem of a given system*, or *consistency of a given system*. Should we really conclude that our conviction that we do have such determinate conceptions ought to depend on the belief that either cosmological hypotheses (A) and (B) or some suitable surrogate for them are correct?

I am not sure what the answer to this is. It would be nice if the mere *possibility* of cosmological hypotheses like (A) and (B) should somehow be enough to ensure the determinacy of 'finite' and thus of f-decidable sentences. But there is a considerable obstacle to seeing how this would go: for instance, we cannot expect a generalization of (*) in which a possibility operator precedes the '$\exists Z$', since then the possibility of infinite sets satisfying Φ would make infinite sets come out finite. (It is *possible* for the universe to be non-standard—in whatever sense it is possible for it to be infinite if in fact it is finite.)

But I do not think it completely obvious that we could not live with the idea that our conception of finitude (and hence of formula, proof, consistency, etc.) is not fully determinate. For even if there is no way to make sense of the idea that 'non-standard' models of 'finite' are 'objectively unintended', still that does not show that 'finite' is not determinate enough to give determinate truth value to typical sentences containing it. So, for any string of less than (say) 10^{20} symbols, it would seem to be a perfectly determinate matter of fact whether it was a formula of a given language; and if it is humanly provable that a certain string is infinite, then that string is definitely not a formula of the language. Perhaps this is all the determinacy that we have reason to be confident of.

But even if one is willing to swallow that our conception of finitude is somewhat indeterminate, one may think that some undecidable sentences must have determinate truth values, contrary to 'extreme anti-objectivism'. In particular, one might think that the Gödel sentence of our fullest mathematical theory must certainly have determinate truth value: it must be determinately true if our fullest mathematical theory is consistent, and determinately false otherwise.

I am going to argue that this argument is question-begging, but before doing so, I should deal with an objection to the way I have formulated it: some may feel that the right way to put what Gödel's theorem shows is that there is no such thing as 'our fullest mathematical theory', or that if there is such a thing, it is not recursively enumerable and therefore has no Gödel sentence. More fully, the view is that no consistent recursively enumerable theory M could be our fullest mathematical theory, since 'our fullest mathematical theory' would have to be closed under Gödelization: if it included M, it would have to include M's Gödel sentence G_M.[18] If we accept this, and exclude the possibility that 'our fullest mathematical theory' is inconsistent, we are left with the conclusion that 'our fullest mathematical theory' either does not exist or is not recursively enumerable.

This objection seems to me misguided: the most one can get from Gödel's theorem is that the phrase 'our fullest mathematical theory' is vague, and that for each consistent and recursively enumerable theory M that is a pretty good candidate for its denotation, $M \cup G_M$ (or a theory that includes that) is also not bad as a candidate for its denotation. Compare 'bald': under the usual crude idealization that baldness depends only on the number of hairs on the head, we know that if

$$\{x : x \text{ has fewer than } n \text{ hairs}\}$$

is a pretty good candidate for the extension of 'bald', then

$$\{x : x \text{ has fewer than } n+1 \text{ hairs}\}$$

is not a bad candidate either. But of course one cannot conclude from this that the extension of 'bald' is closed under the addition of a hair, for that would imply that everyone is bald (as long as they do not have infinitely many hairs). Similarly, the fact that $M \cup \{G_M\}$ is about as good a candidate as M for the denotation of 'our fullest mathematical theory' does not imply that our fullest mathematical theory is closed under Gödelization. And that conclusion would be thoroughly implausible, at least if we stipulate that by 'our fullest mathematical theory' we mean something like 'the set consisting of all our explicit mathematical beliefs, plus perhaps those mathematical sentences we could easily be brought to believe explicitly, plus perhaps their logical consequences'.[19] If we mean anything like this by 'our fullest mathematical theory', then there can be no question that on any way of making the phrase precise, the set exists and is recursively enumerable, and hence (if consistent) is not closed under Gödelization. It is just that there is no maximally inclusive way of making the phrase precise.

How then must the extreme anti-objectivist position be formulated, so as to make explicit the vagueness of 'our fullest mathematical theory'? The position is that in so far as M is a good candidate for our fullest mathematical theory, sentences undecidable in M have no determinate truth value. So if A is a sentence undecidable on all reasonable candidates for our fullest mathematical theory, it definitely has no determinate truth value; whereas if it is decidable on some but not all of the reasonable candidates for our fullest theory, it is indeterminate whether it has determinate truth value. We have a second-order indeterminacy, since 'indeterminate' is tied to 'fullest mathematical theory' (according to the philosophical principles here under consideration) and hence is just as indeterminate as that is.

Let us now return to the argument against extreme anti-objectivism that I claimed was question-begging: the argument, modified slightly to allow for the second-order indeterminacy just noted, was that the Gödel sentence of any candidate M for our fullest mathematical theory must certainly have determinate truth value: it must be determinately true if M is consistent, and determinately false otherwise.

As I have said, this is question-begging. Let us grant for now that, if M is consistent, its Gödel sentence G_M is true, and that, if it is inconsistent, then G_M is false. It only follows that G_M is either determinately true or determinately false if we assume that M is either determinately consistent or determinately inconsistent. And while of course that will be the case if 'finite' is determinate, we are now exploring the possibility that 'finite' is not determinate. Indeed, we are exploring the possibility that the only mathematical claims that are determinately true are those that are provable in M. (And some of those may not be determinately true either, if their negation is also provable in M.) But if CON_M is the standard formalization of the consistency of M, CON_M is not provable in M, unless $\neg\mathrm{CON}_M$ is too; so in so far as the extreme anti-objectivist identifies consistency with the standard formalization thereof, he will say that it is not determinate whether M is consistent and hence it is not determinate what truth value its Gödel sentence has.[20] (Here I have in mind the case where M contains no humanly recognizable inconsistency. If it contains a humanly recognizable inconsistency, it is natural to hold CON_M and G_M determinately false; in a moment I will show that this verdict can be reconciled with extreme anti-objectivism.[21])

I do not deny that it is an awkward feature of extreme anti-objectivism that it may be indeterminate whether a candidate for our fullest theory is consistent (or more accurately, indeterminate whether a candidate for our fullest theory–predicate picks out a consistent theory—see notes 18 and 21). To investigate how serious this awkwardness is, let us look first at what an extreme anti-objectivist should say about the possibility that our fullest theory is definitely inconsistent: say, that it contains a humanly recognizable inconsistency. Probably the best thing to say (as observed in note 7) is that the adoption of an inconsistent theory would make determinately true just those sentences A that are implied by all of the most natural consistent replacements for that theory.[22] Presumably, if a candidate M for our fullest theory contains a humanly recognizable inconsistency, the inconsistency lies outside Peano arithmetic (P), and every natural consistent replacement for M implies both P and CON_P (as does every consistent candidate for our fullest theory). So P and CON_P come out determinately true on all reasonable candidates for our fullest theory, even the definitely inconsistent ones. On the other hand, if M contains a humanly recognizable inconsistency, then presumably every natural consistent replacement of M implies $\neg\mathrm{CON}_M$ (as does every consistent candidate for our fullest theory); so CON_M would then be determinately false. And so too would M itself. These conclusions seem to be what we would intuitively want, so the possibility of definite inconsistency in candidates for our fullest mathematical theory does not seem especially problematic for extreme anti-objectivism.

But what are the consequences for extreme anti-objectivism of its not being determinate whether a candidate M for our fullest theory is consistent? Presumably Peano arithmetic and CON_P still come out determinately true, since I have argued that they come out true whether or not M is consistent; but presumably CON_M and M now lack determinate truth value, since they come out determinately false if M is inconsistent and not determinately false if it is consistent. And these conclusions too seem to be what we ought to expect. So it is hard to see how to reduce to absurdity the position that it is indeterminate whether the candidates for our fullest theory are consistent. I do not claim that that is an attractive position; only that it might be the least unattractive option to adopt if all cosmological hypotheses like those considered in §3 were to prove to be false.

There is another kind of argument against extreme anti-objectivism based on Gödel's theorem. It claims that we know that the Gödel sentences of the candidates for our fullest mathematical theory not merely have determinate truth value, but are true. The argument is a mathematical induction: all the logical and non-logical premises of M are true; the rules of inference preserve truth; so all the theorems must be true; so the theory must be consistent, so the Gödel sentence must be unprovable, and hence true. Of course, the induction cannot be formalized in M; but it is often felt that it is somehow 'informally valid'.[23]

I have doubts about the intelligibility of this idea of 'informal validity', but would rather not rest on them: I will argue rather that the induction is not simply unformalizable, it is fallacious, in that it relies on the incorrect principles about truth that are responsible for the semantic paradoxes. My argument will presuppose that 'true' as applied to mathematical sentences is a perfectly good mathematical notion. Of course, though restrictions of this notion (for example, to sentences that do not themselves contain 'true' and that have only restricted quantifiers) are definable in set theory, the full notion of mathematical truth is not definable in more ordinary mathematical terms, by Tarski's Theorem. Still, it is a notion that we can axiomatize: we all implicitly accept many axioms involving it (for example, the instances of the schema

(T) True('p') if and only if p

for '*p*' not containing the word 'true', and the claim that all of these instances are true); and it is possible to consistently extend these axioms in any of several attractive ways, perhaps adding some special rules of inference involving 'true' as well. Whatever axioms and rules of inference about truth we accept are part of our mathematical theory M.

What I now want to argue is that the inability to carry out formally the inductive argument that all theorems of M are true need not rely on excluding the notion of truth from the formal principle of induction. (If it did rely on that, then, if we could convince ourselves that inductions involving 'true' were 'informally valid', we would have an informal argument for the truth of the Gödel sentence of M.) In addition, the inability to carry out formally the induction need not rest on an inability to pass from asserting of each axiom that it is true to asserting that all of them are true (or analogously for the rules of inference); indeed, there is no difficulty in making such a passage if M has only finitely many axioms and rules, and I know of no strong reason why this should not be so. Rather, the most fundamental source of the problem, I claim, is the assumption for each axiom of M that it is true and for each rule of inference that it preserves truth: some of these assumptions are not only unprovable, but refutable in M.

Exactly which axiom of M fails of truth or which rule fails of truth-preservingness depends on your theory of truth, but in any adequate theory of truth, one of the axioms or rules of M will suffer this fate. For instance, in all versions of the Kripke supervaluational theory (without 'closing off') or of the Gupta–Belnap revision theory,[24] the inference rules include both modus ponens and the inference from A to True('A'). But now consider the claims that these rules are truth-preserving: indeed, consider just the weak schematic forms of these claims, that is, the schemas

I True ('A') & True ('A → B') → True ('B') and

II True ('A') → True('True ('A')').

No version of either the Kripke theory (without closing off) or the Gupta–Belnap theory accepts both these schemas. (The attractive versions of each accept the first schema, but not the second.[25]) And indeed, each version of either theory explicitly rejects certain instances of one or the other schema: adding the schemas not accepted to the theory would engender a version of the Liar paradox. In a Kripke theory obtained by 'closing off' a fixed point, the axiom schema

True ('A') → A

will typically hold; but then the schema

True ('True ('A') → A')

that asserts the truth of all instances of the previous schema has counter-instances (for instance, when you instantiate on the Liar sentence). The point made here for the Kripke and Gupta–Belnap theories can be made with great generality for other theories of truth.[26] The upshot is that the premises of the inductive argument for the truth of all theorems (and hence the consistency of the theory and the truth of the Gödel sentence) cannot all be accepted; the unprovability of the Gödel sentence is not due to an inability to carry out formally an induction that is 'informally valid'.

I have been arguing that the consideration of the Gödel sentence of (a candidate for) our fullest mathematical theory gives no decisive reason for thinking that some undecidable sentences have determinate truth values. But my discussion in the last few paragraphs has implications also in a different context, where we assume (either on account of cosmological assumptions like those in §3, or for other reasons) that the Gödel sentence does have determinate truth value, and we are interested in whether it is determinately true or determinately false. In that context, my claim is that we have less of an argument that the Gödel sentence is true than many people think. Of course, to say this is not to say that we should not hope that the Gödel sentence is true: hoping that is tantamount to hoping our theory is consistent, which seems like a reasonable attitude. It is also not to say that we should not have a positive degree of belief that it is true: we can reasonably have positive degrees of belief in many things that we do not think even informally provable.[27]

This point may be usefully combined with the point, stressed several times, that there is no uniquely best candidate for 'our fullest mathematical theory'. If M is a relatively good candidate for my fullest mathematical theory, I am likely to have a fairly decent degree of belief in G_M. This will probably make $M' = M \cup \{G_M\}$ a pretty good candidate for my fullest mathematical theory too, given the way my degrees of belief work, even though G_M is not even informally provable from M. (The theory M' may be a better candidate than M is, if M is 'on the weak side' of the cluster of theories that are candidates for my fullest mathematical theory.) We can reiterate the addition of a Gödel sentence through the constructive ordinals, as in Feferman (1962). My position is that at some point in this process, the claim of the theories to be 'our fullest mathematical theory' begins to decrease gradually. It decreases gradually, as opposed to dropping off suddenly: this seems to me the natural attitude to take, and it is one that is facilitated by taking the considerations that favor passing from a theory to its Gödel sentence as less than 'informal proof'.

These last two paragraphs have been a bit of a digression. The main theme of this section is that the consideration of Gödel's theorem gives no decisive reason for thinking that some undecidable sentences have determinate truth value. And Putnam's argument makes it hard to see how any sentences undecidable in all candidates for our fullest theory could have determinate truth value, if cosmological hypotheses like those considered in §3 are incorrect.

Notes

1. An interesting recent paper relevant to its defense is (Hawthorne 1996).

2. I do not mean to commit the standard platonist to assuming that 'domains' or 'structures' are sets; the view that set theory is true of some determinate domain is not supposed to involve a commitment to a set of all sets.

3. Mark Balaguer (1995) calls it 'full-blooded platonism'. An earlier version of Balaguer's paper (a chapter from his dissertation) helped convert me to the attractions of the position. (I had hinted at something like it earlier, but wihout following through on its implications: see (Field 1989), pp. 275–8, especially the discussion of whether there is a broadest possible notion of set.) David Papineau (1993) calls the position 'postulationism'; he argues against it. My term 'plenitudinous platonism' is an adaptation of the term 'plentiful platonism' in Penelope Maddy's chapter (1998*c*).

4. More generally, one might hold that whenever you have a theory that postulates mathematical objects, then, as long as the theory is 'nominalistically correct', in the sense that all its consequences that do not postulate such objects and do not use specially mathematical vocabulary are correct, there are mathematical objects that satisfy the theory under a perfectly standard satisfaction relation.

5. For a response to one doubt about the coherence of plenitudinous platonism, see (Field 1994, Appendix). (Many of the other main ideas of the present chapter are discussed there as well.) Tony Martin raised another interesting doubt after the conference which I do not have the space to pursue here.

6. Quine pointed out long ago that there is no hope of making sense of the idea that mathematics and logic together are true by convention (since logic is required in determining the consequences of one's conventions); but the difficulty does not arise if one confines the claim to mathematics alone.

7. I am assuming in this paragraph that the mathematics we accept is consistent: were we to accept inconsistent mathematical claims, we would not want to take them all as true. (Probably the best thing to say is that, if our overall mathematical theory is inconsistent, then anything implied by all the most natural consistent replacements of it is true by convention, whereas anything implied by some, but not all, of the natural consistent replacements gets no determinate truth value from our inconsistent convention.)

(I take consistency as a logical notion rather than a mathematical one (see (Field 1991)), so that taking our mathematics as true merely by convention does not make consistency claims true merely by convention. I do not doubt that our mathematical conventions could indirectly affect the notion of consistency; but they do not do so in a way that makes it trivial that our mathematical conventions come out consistent.)

8. By a 'mathematical sentence' I do not mean any sentence that mentions mathematical objects or employs mathematical vocabulary: that would allow mathematically formulated sentences of physics to count as mathematical, and I do not intend (DTV) to cover them. On the other hand, 'sentence that neither postulates non-mathematical objects nor employs non-mathematical vocabulary' is too restrictive: it is a reasonable definition of a pure mathematical sentence, but I think 'There is a set of all grapes' should count as (impurely) mathematical. Roughly, by a mathematical sentence I mean the sort of sentence that would be appropriate to be settled by mathematical considerations alone. But it is too much trouble to try to make this precise, and for present purposes it will not do much harm to restrict attention to pure mathematical sentences.

9. A simple example of how this could happen in principle is the following: suppose that our fullest mathematical theory did not imply the axiom of infinity (taken as the axiom that there are infinite pure sets), but did include both the axiom that there is a set of all physical things and a replacement schema that applies to physical vocabulary as well as mathematical. If the determinate truths of physics include the claim that there are infinitely many physical things discretely ordered under some physical relation, then the axiom of infinity follows from true physics plus the mathematics, but not from the mathematics alone. (The point is that set theory with the denial of the axiom of infinity is not 'conservative': see (Field 1989, pp. 56–7), for an elaboration.) A similar illustration might in principle arise for the axiom of inaccessibles, but only if the claim that there are inaccessibly many physical objects could in principle achieve the status of a determinate truth about the physical world.

10. Assuming it to be consistent.

11. This is so even if consequence and consistency are not defined in terms of proof (or in terms of model either), but are related to proof and model indirectly by a 'squeezing argument': see (Field 1991).

12. So attempts to use the second-order consequence relation to evade Putnam's argument are question-begging. For a further discussion, see (Weston 1976) and (Field 1994).

13. See, for instance, (Lewis 1984). That reading is not without textual support, both from other writings of Putnam at about the same time as (Putnam 1980), and even from the second half of that paper; nonetheless it seems to me to miss the interesting argument in the first half of that paper.

14. By 'infinite set' here I really mean 'object y such that for infinitely many x, $\langle x, y\rangle$ satisfies '$\in$' in the interpretation'.

15. If we modified (ii) in $\Phi(Z)$ to say merely that there is some unit of time such that any two members of Z occur at least that unit of time apart, the infinitude assumption becomes easier to satisfy (though still non-trivial); but the Archimedean assumption would be correspondingly harder to satisfy, given the possibility of points that are infinitesimally close. Since it might in any case be doubted that there is a definite physical fact as to whether there

are infinitesimals, I think it safer to stick with a formulation for which they are irrelevant.

16. Although the argument as presented assumes the complete determinacy of the physical vocabulary (relative to a choice of domain of quantification), I think it is clear that relaxing this would not undermine the force of the argument.

17. Given a reasonable assumption about the limits on how determinate quantification (even restricted to physical things) can be, this follows from a slight generalization of the usual relative consistency proofs to the case of impure set theory, coupled with a downward Skolem–Löwenheim argument. But, even if the assumption about the indeterminacy of quantification is relaxed, the prospects for giving sentences about the size of the continuum definite truth value are quite bleak. There is no space for discussion of these points, but the discussion on pp. 415–16 and 419–20 of (Field 1994) (in addition to that in (Putnam 1980)) might provide enough to enable the reader to extrapolate them.

18. Strictly speaking, 'the Gödel sentence of M' is ill-defined: the Gödel sentence depends not just on M but on a way of defining 'axiom of M' and 'rule of inference of M' (as well as on various choices that can be made once for all theories in a given language, such as a Gödel numbering). To set things right, the discussion in the text (here and in subsequent paragraphs) should really be done not in terms of (candidates for) our fullest mathematical theory M, but in terms of (candidates for) our fullest theory–predicate S: where a theory predicate is an RE-formula (in one variable) which numerates the axioms and rules of some theory M. (I am using the terminology of (Feferman 1962).) For instance, the sentence to which this note is attached should read: the view is that no theory–predicate S that numerates a consistent mathematical theory could be our fullest mathematical theory–predicate, since 'our fullest mathematical theory–predicate' would have to be closed under Gödelization: if what it numerates included S, it would have to include S's Gödel sentence G_S.

(The correction here is of little significance for finitely axiomatized theories, but is of more importance when M is not finitely axiomatized. And it is arguable that in that case, theory–predicates rather than theories are more directly 'psychologically real': that the theory predicate is the means by which an infinitely axiomatized theory is represented in our finite heads.)

For readability I have chosen to speak sloppily of theories instead of theory–predicates in the text, though in a note I will mark one place where it is of philosophical interest to observe the more correct formulation.

19. We would probably want to refine this a bit, to reduce the chance that all reasonable candidates for our fullest theory are inconsistent; but this will not affect the case for recursive enumerability.

20. On the other hand, if T is a theory that (like Peano arithmetic) is provably consistent in all reasonable candidates for our fullest theory, then G_T and CON_T will be determinately true.

21. As I observed in note 18, the sort of formulation given in this paragraph

is rather sloppy: to set things right, I should speak not of a theory M, but of a theory–predicate S that picks out a theory. This is of philosophical interest in the present context, since it points up the fact that, intuitively, the indeterminacy as to the truth of G_S and CON_S may be partly due to the fact that when 'finite' is not assumed determinate, there can be an indeterminacy in which theory the predicate S picks out.

22. Anyone who thinks that my use of 'inconsistent' and 'consistent' here assumes the determinacy of the notions should recall the second paragraph of §2. But if you like, you can replace 'inconsistent' by 'definitely inconsistent' here, and likewise for 'consistent'. Similar remarks apply to much of the discussion in the next few paragraphs.

23. The reader will note that this argument would, if valid, establish only the truth of the Gödel sentence, not its determinate truth. The gap could be filled by the principle that, if we can informally prove the truth of something, it must be determinately true (a principle that is plausible, though not beyond controversy); alternatively, we could have done the induction in terms of determinate truth in the first place (though the premises of the induction then seem somewhat less obvious). Either way, the argument for the determinate truth of the Gödel sentence is a bit shakier than the argument for its truth. But I want to respond to even the less shaky argument.

24. Kripke and Gupta–Belnap present their accounts as explicit definitions of truth; to do this requires that the quantifiers of the object language be interpreted as ranging over less than everything (in particular, ranging over some set D), so what they are really defining is not truth but 'comes out true when the quantifiers are restricted to domain D'. Obviously we could not use such restricted truth predicates in the inductive argument under consideration. But the Kripke and Gupta–Belnap definitions are of interest in that one can investigate which principles of truth they validate, and one can then use those principles in an axiomatic theory of truth; that way the restriction to a domain is not required, and the use of the truth predicate in the inductive argument is *prima facie* more promising. This approach (advocated in (McGee 1991)) is what I am presupposing.

25. By a 'version' of the Kripke theory or the Gupta–Belnap theory, I mean a decision as to what conditions we impose on: (i) the sets of sentences we supervaluate over in the Kripke case; (ii) the sets of sentences we allow in the limit stages in the Gupta–Belnap case. Attractive versions of these theories impose closure under modus ponens as one of these conditions; this guarantees Schema I, but at the same time rules out Schema II.

26. For more on this, see (Field 1994, note 18).

27. It also is not to deny that there is something awkward about believing M while simultaneously believing $\neg G_M$: this is tantamount to believing M and at the same time believing $\neg\mathrm{CON}_M$, and (at least in the context of an assumption of the determinacy of finiteness, where there is no question that $\neg\mathrm{CON}_M$ adequately formalizes the consistency of M) the beliefs do not cohere

well with each other. So, were we to believe both M and $\neg G_M$, we would have motivation to revise one of our beliefs; but it could be M rather than $\neg G_M$ that was the prime candidate for revision, so it is hard to see how this consideration gives reason to believe G_M.

Bibliography

Balaguer, M. (1995). A Platonist epistemology. *Synthèse*, **103**, 303–25.

Boolos, G. (1984). To be is to be the value of a variable (or some values of some variables). *Journal of Philosophy*, **81**, 430–9.

Feferman, S. (1962). Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, **27**, 259–316.

Field, H. (1989). *Realism, mathematics and modality.* Blackwells, Oxford.

Field, H. (1991). Metalogic and modality. *Philosophical Studies*, **62**, 1–22.

Field, H. (1994). Are our logical and mathematical concepts highly indeterminate? In *Midwest studies in philosophy* (ed. P. A. French, T. E. Uehling, Jr., and H. K. Wettstein), Vol. 19, pp. 391–429.

Gupta, A. and Belnap, N. (1993). *The revision theory of truth.* MIT Press, Cambridge, Massachusetts.

Hawthorne, J. (1996). Mathematical instrumentalism meets the conjunction objection. *Journal of Philosophical Logic*, **25**, 363-97.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690-716.

Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, **62**, 221–36.

Maddy, P. (1998*c*). How to be a naturalist about mathematics. *This volume*, 161–80.

McGee, V. (1991). *Truth, vagueness and paradox.* Hackett Publishing Co., Indianapolis.

Papineau, D. (1993). *Philosophical naturalism.* Blackwells, Oxford.

Putnam, H. (1980). Models and reality. *Journal of Symbolic Logic*, **45**, 464–82.

Weston, T. (1976). Kreisel, the continuum hypothesis, and second order set theory. *Journal of Philosophical Logic*, **5**, 281–98.

Department of Philosophy
New York University
100 Washington Square East
New York, NY 10003
USA
email: hf18@is4.nyu.edu

17

Two conceptions of natural number

Alexander George and Daniel J. Velleman

The distinction between the completed and the potential infinite is well known. Less noted is a corresponding contrast between two different conceptions of natural number. It is only to be expected that there would be such a contrast, since the natural numbers form our most basic model of an infinite collection. In this note, we present these two distinct conceptions by articulating the philosophical visions that inspire them and the mathematical definitions that give them substance. We show how these analyses satisfy, in interestingly different ways, the basic demands that any such definition must meet. In keeping with the fundamental difference in perspective between these accounts of natural number, we should expect that those who advance the one definition will find the other wanting. We try to describe what form these respective criticisms will take and to say why they will appear misguided to proponents of the conception against which they are directed.

An intuitive way to try to characterize the natural numbers is to use the mathematical idea of the closure of a set under an operation. If A is a set and f is an operation, then A is said to be *closed* under f if, for every object a in A, the result of applying the operation f to a, denoted $f(a)$, is also in A. The *closure* of A under f is the smallest set containing A that is closed under f. For example, in this terminology the set $\mathbb{N}$ of natural numbers would be the closure of the set $\{0\}$ under the successor operation S.[1]

There are two ways that mathematicians commonly form the closure of a set A under an operation f. The first is to begin with the set A, and add additional elements to form the closure. For example, if a is in A, then $f(a)$ must be added to A if we are to obtain a set that is closed under f. But then $f(f(a))$ must also be added, and then $f(f(f(a)))$. In fact, anything that can be obtained by applying the operation f repeatedly to elements of A must be in the closure of A under f. Let A_* be the set of all elements of A, together with those objects obtainable by applying f repeatedly to elements of A. Then A_* is closed under f, and therefore it is the closure of A under f.

Another way to form the closure of A under f is to let A^* be the intersection of all sets that contain A and are closed under f. In other words, the elements of A^* are those objects that have the property of belonging to every set that contains A and is closed under f. It is not hard to see that A^* contains A and is closed under f, and as the intersection of all sets with this property it must be the smallest such set. Thus A^* is also the closure of A under f, and therefore $A^* = A_*$. (See, for example, (Enderton 1972, pp. 22–5).)

These two ways of forming the closure of a set under an operation suggest two ways of trying to characterize the natural numbers. If we let $A = \{0\}$ and f be the successor operation, then they correspond to the definitions we have given of A_* and A^*. The first characterization says that the natural numbers are just those objects obtainable from 0 by repeatedly applying the successor operation. We might regard this characterization as giving two rules for generating natural numbers:

(1) 0 is a natural number, and

(2) If n is a natural number, then so is $S(n)$.

The natural numbers, according to this characterization, are those, and only those, objects that are generated by these rules, so it is natural to call it the *build up* (BU) definition of "$\mathbb{N}$". The restriction that only objects generated by rules (1) and (2) are numbers is often referred to as the *extremal clause* of the definition.

The second characterization says that the natural numbers are precisely those objects that belong to every set that contains 0 and is closed under the successor operation. This characterization starts with sets that contain 0 and are closed under successor, most of which are larger than the set of natural numbers, and then eliminates the non-numbers from these sets by intersecting them. It is therefore appropriate to call it the *pare down* (PD) approach to defining the natural numbers.

The pare down definition of the natural numbers was first advanced independently by Richard Dedekind and Gottlob Frege in the nineteenth century. Note that the validity of mathematical induction is easily seen to follow from the PD definition. For if a predicate holds of 0, and it holds of $S(n)$ whenever it holds of n, then its extension is a set containing 0 and closed under S. Since any natural number, according to the PD definition, must belong to every set containing 0 and closed under S, it follows that the predicate in question must apply to every natural number.

It is also apparent that the PD definition captures all and only the natural numbers. "All" because each natural number belongs to every set that contains 0 and is closed under the successor operation; and "only" because anything that belongs to every such set will also belong to $\mathbb{N}$, since $\mathbb{N}$ is itself such a set. But note that this reasoning will apply only if we reckon the set of natural numbers to be in the range of the second-order quantifier in the PD definition. This impredicativity may lead to concern. If one views the definition as offering a recipe for constructing the set of natural numbers, impredicativity is fatal: for it would then have us creating the collection of natural numbers through appeal to that very collection.

But this is clearly not the purpose of the definition according to those who offer it. Its object is not to create the collection of natural numbers, which, on the contrary, is viewed as already existing, but rather to identify which of the existing completed collections of objects ℕ is. Viewed as picking out the already existing collection ℕ, it is, as W.V. Quine has remarked, "not visibly more vicious than singling out an individual as the most typical Yale man on the basis of averages of Yale scores including his own" (1969, p. 243).

It should be clear that the PD definition of "ℕ" is most natural from a platonist perspective. For the PD approach is at home with a conception of the natural numbers as a completed infinite collection that exists, independently of our activity, amidst other likewise completed infinite sets. On this view, the task of an adequate definition is to pick out the set ℕ from this universe of entities.[2]

But now consider the matter from a constructivist standpoint. There is no circumscribed domain of sets "out there" in advance of our activity. The domain of sets is indefinitely extensible; that is, given any particular delimitation of the universe of sets, we can construct in terms of it another set not previously in the universe. In particular, the set of natural numbers will not exist until we construct it. Now the impredicativity of the PD definition is a serious problem. For given this definition, how can one show that, to cite a famous example, Julius Caesar is not a natural number? Only if there exists a set containing 0 and closed under successor that does not contain Caesar. The argument that ℕ itself is such a set will not satisfy the constructivist, since ℕ cannot be assumed to exist before it is constructed.

The impredicativity of the PD definition means that it cannot be viewed as a procedure for the construction of ℕ. The build up definition, however, can be. The BU approach coheres best with a constructivist stance, according to which a definition of "ℕ" should provide us with an account of how to generate all and only the natural numbers. For the constructivist, the only exclusionary clause that is required is one to the effect that only those objects generated by the two specified rules of construction are natural numbers. Once we know this, we can see, for example, that Julius Caesar is not a natural number, for he is not identical to the output of either rule.[3]

But just as the PD approach appears problematic from the constructivist perspective, so does the BU approach appear wanting to the platonist. Yes, the platonist will grant, the BU definition excludes the use of other than the intended rules in generating the members of ℕ, but it does not explicitly exclude unintended uses of those rules. In particular, there is nothing in the BU definition that bars non-finite iteration of the second generating rule. And such iteration must be ruled out, the objection continues, for otherwise there is no guarantee that the definition will capture only the natural numbers, and nothing else.[4] The complaint is not that those who offer the BU definition fail to realize that only finite iteration is permitted, but rather that this realization is no thanks to the definition.

For this reason, the BU definition will appear at best elliptical from the platonist perspective: it must be understood that the second rule of the definition permits only *finite* iteration of the successor operation to yield natural numbers. Of course, if "finite iteration" means "iteration n times, for some natural number n," then the definition is circular, as Dedekind himself objected:

> If one presupposes knowledge of the sequence N of natural numbers and, accordingly, allows himself the use of the language of arithmetic, then, of course, he has an easy time of it. He need only say: an element n belongs to the sequence N if and only if, starting with the element 1 and counting on and on steadfastly, that is, going through a finite number of iterations of the mapping φ [...], I actually reach the element n at some time; by this procedure, however, I shall never reach an element t outside of the sequence N. But this way of characterizing the distinction between those elements t that are to be ejected [...] and those elements n that alone are to remain is surely quite useless for our purpose; it would, after all, contain the most pernicious and obvious kind of vicious circle. The mere words "finally get there at some time," of course, will not do either; they would be of no more use than, say, the words "karam sipo tatura," which I invent at this instant without giving them any clearly defined meaning. (Dedekind 1967, pp. 100–1)[5]

Consequently, the platonist might offer the following as a friendly amendment to the BU proposal: delete the extremal clause, which perforce will be either inexplicit or circular, and secure its intent by specifying that induction is valid. For the platonist, the requirement that induction is a valid means of forming generalizations about the elements of some collection guarantees that non-numbers will be excluded from the collection. For the predicate "natural number" applies to 0, and applies to $S(n)$ if it applies to n, and therefore it must apply to every element of a collection for which induction is valid. Of course, this entailment presupposes that "natural number" (or a predicate of a piece with it) is taken to be well-defined: the validity of induction does not articulate the intentions of the extremal clause unless induction is understood impredicatively, as a generalization over a pre-existing domain of predicates that includes the very predicate being defined. As we have observed, this impredicativity is unacceptable from the constructivist point of view, and therefore the constructivist will not consider induction to be an adequate replacement for the extremal clause.[6, 7]

In fact, the constructivist will find the friendly amendment not only unhelpful, but unnecessary as well. For just as the constructivist's objection to the PD approach appears off the mark to the platonist, so too will the objection about BU's inexplicitness seem to the constructivist. From the constructivist viewpoint, no intelligible but unwanted possibility has yet been described that would require changes to, or replacement of, the extremal clause: since the constructivist accepts the potential infinite, but not the completed infinite, the idea of applying the generation rules an infinite number of times is unintelligible from the constructivist perspective, and so nothing need be said to rule out those alleged entities that would be constructed as a result of such an impossible application.[8, 9, 10]

If BU is not to postulate induction, how then is its validity to be secured?

The argument traditionally offered by constructivists is just this. Consider a predicate P for which the premises of induction hold and let n be any given natural number. The second premise of induction tells us that if P holds of 0 then P holds of $S(0)$. Taken together with the first, which states that P does hold of 0, we can conclude that P holds of $S(0)$, by modus ponens. Since we know, again by the second premise, that if P holds of $S(0)$ then it holds also of $S(S(0))$, we can likewise infer that P holds of $S(S(0))$. And so on. Thus, we see that at every stage of a construction that begins with 0 and proceeds by repeatedly applying the successor operation, P must hold of the object constructed. But on the BU view, n was obtained by precisely such a construction (this is what the extremal clause asserts), so we may conclude that P holds of n. Thus, induction is valid with respect to any well-defined predicate.[11]

Earlier, we noted that the impredicativity of the PD definition of natural number renders it unacceptable to the constructivist, who will turn instead to the BU account for an adequate analysis. This turn will clearly be rewarding for those who believe that in assessing a definition's impredicativity, it suffices to confine one's examination to that definition. For, as we have seen, the source of the impredicativity is the second-order quantifier in the principle of mathematical induction, and induction is not actually a part of the BU definition, but is rather a consequence of it.

But someone might think that the assessment of a definition's impredicativity cannot proceed through scrutiny of it alone, but also requires examination of conceptual truths about the defined notion, in this case the validity of induction.[12] Or someone might be convinced that a definition of "$\mathbb{N}$" cannot merely specify the extension of "natural number" but must also specify the grounds for generalizations about natural numbers; even if induction is not used for the first task (say, by appealing to the BU definition), it is needed for the second, and hence reference to it will have to be made in any adequate analysis of the meaning of "natural number."[13] While we do not here endorse these, or similar, proposals, they do make it worthwhile to inquire whether impredicativity lurks in the BU account of induction and, if it does, whether it is of the same nature as that encountered in the PD conception.

In fact, there is something in that account about which one might worry. According to the constructivist, the domain of predicates to which induction applies is indefinitely extensible. Indeed, the very act of defining "$\mathbb{N}$" extends this domain, by creating the new predicate "natural number," along with other predicates defined in terms of it. Induction ought to apply to these new predicates. We have argued that the BU definition implies the validity of induction, but might it imply it only for predicates previously defined? If so, then the BU definition would in a sense undermine itself: it would justify induction for all previously defined predicates, and it would simultaneously create new predicates, thus rendering obsolete the very version of induction that it implies.

This same self-sabotage is avoided in the PD definition at precisely the cost of impredicativity. For this definition picks out a set $\mathbb{N}$, which then gives us the new predicate "natural number" and other predicates defined in terms of it, and these new predicates have the potential to undermine the work done previously by the PD definition, since their extensions should have been included among the collections that were intersected to produce $\mathbb{N}$ in the first place. Of course, from the platonist perspective, they were indeed included, since they existed all along, and so all is well. But a constructivist cannot likewise argue that the BU justification of induction goes through because all well-defined predicates already exist quite independently of mathematical activity, for this is precisely what a constructivist denies. How, then, to respond to this concern about BU's justification of induction?

We believe that Charles Parsons suggests the answer when he notes that:

> the principle [of induction] refers to arbitrary predicates, without any assumptions having been made about what counts as a predicate. Like the principles of predicate logic itself, we have a purely formal generalization about predicates, which is not a generalization over a given *domain* of entities and could not be, since it is not determined what predicates will or can be constructed and understood. (Parsons 1992, p. 143)

Our understanding of Parsons' insight is as follows. Often when one makes a generalization—say, that everything in some domain D has property P—one justifies it by examining the objects in D. The most straightforward case would be when one examines the elements of D one by one to see if they have property P (as may happen when the domain D is finite). If we take this as our model, we might be inclined to say that one cannot arrive at a generalization about the elements of a domain D until one knows what is in the domain. But this is not always true. The intuitionistic understanding of "every" illustrates another possibility: even if D is indefinitely extensible, one might arrive at the generalization that everything in D has property P by examining P, not the elements of D, and realizing that something about P makes it true of anything that we would allow into D, even if we do not yet know what is in D. This is precisely what happens with induction, on the BU view. According to the latter, we believe induction applies to all predicates, not because we have surveyed the available predicates and noted that the induction principle applies to all of them (this procedure would indeed lead to the feared self-undermining), but rather, as we saw above, because of the extremal clause, which implies that induction will apply to any predicate, even predicates not yet constructed. Thus, one should not suppose that a grasp of the whole domain of predicates is needed in order to understand why induction holds for all predicates. If this justification still involves impredicativity, then it is an attenuated impredicativity that should be distinguished from that present in the PD approach.

It may well be that an analysis of natural number that succeeds in justifying the principle of induction will have either to be impredicative or to interpret constructively the principle's second-order quantifier. Someone who rejects these two options will find it difficult not to reject the induction principle in its full generality.[14] In any event, one advantage of sharply distinguishing, as we have, between the two accounts of natural number is that doing so enables one to see that the oft-repeated claim that all definitions of natural number are impredicative elides interesting differences.[15]

We would like to conclude, however, by stressing an important affinity between these two approaches, one that has perhaps not always been recognized. In the case of both the PD and the BU definitions of natural number, something has to be in place in order for them to be taken as intended by someone trying to learn the meaning of "natural number." In the first case, the learner must understand the second-order quantifier as ranging over a pre-existing completed totality of sets, including the set $\mathbb{N}$ itself. And in the second, the learner must grasp the concept of finite iteration (perhaps because, as we saw constructivists would insist, this is the only kind of iteration that is intelligible to him). Without this stage-setting, the definitions cannot be understood as intended by their respective proponents.

It is important to notice that what must be in place is, in each case, akin to a grasp of the very notion being defined. These definitions are not circular, but taking them in the intended ways does presuppose some understanding of the very concepts being defined. Let us call such definitions *elucidations.*[16]

We do not want to say that elucidations must fail to convey the target concepts to someone who does not already possess the relevant understanding. After all, the BU and PD elucidations, as a matter of fact, often do help students to acquire the defined notions. Rather, our point is that such elucidations cannot convey these notions to someone who lacks them *through being understood as intended*, for these definitions cannot be so understood except by someone who already grasps in essence the notions being defined.

Elucidations that succeed in conveying an understanding of a term are comparable to speech to infants that facilitates acquisition of language. For such talk likewise does not convey knowledge of language through its being understood as intended—if it were so understood, there would be no need to convey this knowledge.

We will not speculate here regarding how such learning is accomplished. We do not know of any reason, though, for distinguishing between the BU and PD conceptions as regards their conveyability. In the first place, the conceptions behind the BU and the PD definitions are, of course, both infinitistic, and as such neither can be exhaustively displayed by any observable stretch of human behavior.

Secondly, if a learner lacks the relevant conceptions (say because they are not given as part of his innate conceptual endowment), then, as just noted, he cannot gain either of them by taking in their definitions as intended, for so understanding them requires a grasp of something akin to those very conceptions. And if under these circumstances a learner can nevertheless somehow work his way to the target conceptions by taking in their definitions as other than intended, then, pending further information regarding how this takes place, we cannot say that the one conception is any more easily arrived at than the other.

Finally, if appeal to innate notions is made, there is no *a priori* reason why the one conception should be natively given to us and not the other. It might be tempting to argue that there is such a reason, namely that the PD conception is not intelligible and so not there to be given to us, whereas the BU conception is. But if the argument for unintelligibility is ultimately grounded (as it often is) on considerations of acquisition, we are plainly moving in a circle.

In conclusion, though we have been at pains to show that the above two conceptions of natural number do indeed differ in significant ways, we cannot say with any confidence that they do so in point of conveyability.[17]

Notes

1. We are not concerned here about exactly how 0 and the successor operation are defined. One could, for example, use von Neumann's set-theoretic definitions $0 = \emptyset$ and $S(x) = x \cup \{x\}$, or one could regard numbers as strings of strokes and take 0 to be the empty string and the successor operation to be the operation of adding one more stroke to a string. The discussion in the rest of the paper would apply equally to either definition.

2. Henri Poincaré seems to have been the first to note the link between a PD approach to the natural numbers and a commitment to the completed infinite; see sections VIII and XI of "The Last Efforts of the Logisticians" in (Poincaré 1952). Because Poincaré held that "*There is no actual infinity*" (p. 195, original italics), he also rejected PD-type definitions of the natural numbers.

3. Although we have emphasized the strong connection between, on the one hand, PD and platonism, and, on the other, BU and constructivism, note that we have taken no position here regarding whether PD requires a platonist perspective, or BU a constructivist one.

4. For example, if numbers are taken to be strings of strokes, we must ensure that infinitely long strings of strokes are excluded from $\mathbb{N}$.

5. The circularity becomes even more apparent if we try to formalize in set theory the build-up method of forming the closure of a set A under an operation f. The usual approach is to define recursively a sequence of sets A_0, A_1, A_2, ... by letting $A_0 = A$ and, for each n, $A_{n+1} = \{ f(x) : x \in A_n \}$. The set A_* can then be defined to be the union of all sets A_n. Of course, our sequence of sets is indexed by the natural numbers, so it would be circular to use this method (with $A = \{0\}$ and $f = S$) to define the natural numbers.

6. Poincaré may be the earliest proponent of the BU approach conscious of the distinction between it and the PD perspective. According to one definition, Poincaré says, "*a finite whole number is that which can be obtained by successive additions, and which is such that n is not equal to n* − 1," while the other holds that, as he puts it, "*a whole number is that about which we can reason by recurrence.*" Poincaré continues: "The two definitions are not identical. They are equivalent, no doubt, but they are so by virtue of an *a priori* synthetic judgment; we cannot pass from one to the other by purely logical processes" ("The New Logics," reprinted in (Poincaré 1952, p. 173), original italics). The "*a priori*

synthetic judgment" here is the validity of mathematical induction, for, as we shall see shortly, inferring that induction holds from the BU definition requires the use of induction itself. Because the second definition of natural number (essentially, the PD account) is unacceptable to Poincaré (see note 2 above), he concludes that the validity of mathematical induction cannot be established by purely logical means from any adequate account of natural number and, hence, that the logicist reduction of arithmetic fails.

7. Someone disturbed by a perceived inexplicitness in the BU definition might alternatively offer the following amendment: specify that when the second generation rule is iterated, the set of steps in the iteration must be Dedekind finite, where a set is said to be *Dedekind finite* just in case it cannot be mapped one to one onto any proper subset of itself. The amended definition is not circular, for it does not employ the notion of "finite" or "natural number," and it avoids the use of induction in securing the effect of BU's extremal clause. Yet, this proposal would likewise be rejected by a proponent of BU for it continues to involve an impredicativity. To say that a set is Dedekind finite is to say that there does not exist a function of a certain kind. This claim therefore involves quantification over all functions, including those defined in terms of the natural numbers. (For example, this is the reason why Solomon Feferman and Geoffrey Hellman (1995) chose not to take this approach; see their note 3 and page 15.)

8. Michael Dummett, for example, seems to suggest that no replacement for the extremal clause is needed:

> Even if we can give no formal characterisation which will definitely exclude all such elements, it is evident that there is not in fact any possibility of anyone's taking any object, not described (directly or indirectly) as attainable from 0 by iteration of the successor operation, to be a natural number. (Dummett 1978, p. 193)

9. Even those who accept the completed infinite can defend the extremal clause of the BU definition against the criticisms of the platonist, if they are willing to accept the concept of finiteness as being understood in advance of the characterization of the natural numbers, and then to use this concept to express the extremal clause. This seems to be the standpoint taken by Feferman and Hellman (1995). Their approach, in effect, is to prove the existence of structures satisfying Peano's axioms by constructing an example of one. The universe of their example is defined to be the set of all those objects x such that there exists a finite set containing precisely the predecessors of x under iteration of the operation S, with 0 being the only one of these predecessors that does not itself have a predecessor. This finite set could be thought of as recording the process of constructing x by a finite iteration of the successor operation, beginning with 0, and thus this definition could be thought of as a version of the BU definition. Note that it is important that the recording set be finite so as to ensure that the iteration is finite. (For their analysis, see the first line of their proof of Theorem 7 on p. 10, and their definition of "Fin" on p. 4. The requirement that the recording set be finite is enforced in their formal system by the axiom

(Card), which guarantees that the set is Dedekind finite. This axiom is used to prove that the induction axiom holds in their example. For some discussion of the history of this proposal, and further elaboration on the relationship between their definition and the BU definition, see the following note.)

Feferman and Hellman call their approach "predicativism", or "predicative logicism", and contrast it with classical logicism as follows:

> Classical logicism provides a complete analysis of the concepts "finite", "infinite", and "cardinal number", but at the price of *impredicative comprehension* with all of its attendant "metaphysical" commitments. Predicativism avoids the latter but must presuppose the concept of "finite" in some form or other. However, [...] it can do this in a natural way *without thereby taking the natural number system as given.* (p. 15)

The fact that predicativism must presuppose the concept "finite" will make it unacceptable to anyone who believes that this concept is as much in need of analysis as the concept "natural number". As Daniel Isaacson (1987) suggests, the predicativist definition will be successful only if (i) the second-order quantifier in the definition ranges over a domain that includes all finite initial segments of $\mathbb{N}$, and (ii) the domain contains no infinite sets. He concludes that the definition therefore "does not fare significantly better on the score of avoiding impredicativity than the one based on full second-order logic" (p. 156). Feferman and Hellman (1995, note 5, p. 16) argue in response that the existence of the required finite initial segments can be justified predicatively, but it seems to us that they have failed to answer part (ii) of Isaacson's objection, namely that infinite sets must be excluded from the domain of quantification. As we saw earlier, it is this exclusion of infinite sets from the second-order domain that guarantees that Feferman and Hellman's definition will capture *only* natural numbers. In fact the difficulty here is in effect the same as the difficulty that the platonist finds with the BU definition; it is not the inclusion of the desired elements in the domain that causes problems, but rather the exclusion of unwanted elements.

Charles Parsons also considers a similar definition and finds it wanting for the same reason:

> As a defense of the claim that induction on natural numbers is after all predicative, this exercise is hardly impressive. What has been assumed about finite sets will just reinforce the reply that although perhaps one can escape the impredicativity of induction on natural numbers, one merely throws the matter back to the notion of finite set, where the same problems will arise. (Parsons 1992, p. 148)

10. Feferman and Hellman (1995, note 5), say that their approach "realizes in effect a suggestion attributed to Michael Dummett". This appears to raise a problem for our analysis, according to which Feferman and Hellman's proposal is to be reckoned a BU approach to the natural numbers, for Dummett's suggestion was originally attributed to him by Hao Wang, who claims that it "is more

closely related to the Frege–Dedekind definition [than to the approach of Zermelo, Grelling, and Bernays, who manage without the axiom of infinity]" (Wang 1963, p. 52).

Illumination of this apparent conflict is not furthered by the variation one finds in descriptions of Dummett's suggestion. Wang attributes to Dummett a definition of "ℕ" according to which $k \in \mathbb{N}$ just in case

(i) $(\forall X)((0 \in X \,\&\, (\forall y)((y \in X \,\&\, y \neq k) \rightarrow S(y) \in X)) \rightarrow k \in X)$ and
(ii) $(\exists X)(0 \in X \,\&\, (\forall y)((y \in X \,\&\, y \neq k) \rightarrow S(y) \in X))$.

Parsons, referring to Wang's attribution to Dummett, offers a definition similar to (i) and (ii) and traces the idea back to Zermelo and Grelling (Parsons 1987, p. 206). Isaacson, by contrast, though likewise referring to Wang's attribution to Dummett, offers only clause (i). He adds, however, that in order for (i) "to define anything" (ii) must also obtain for every k (Isaacson 1987, p. 155). Feferman and Hellman (1995, note 5) apparently following Isaacson, also give only clause (i) when describing Dummett's definition. In spite of this, their actual definition is closer to (ii) than to (i), being essentially existential rather than universal.

This confusing variation might be due to different assumptions about the range of the definition's second-order quantifier. If it is assumed to include infinite sets, then (ii) alone will not suffice to exclude all non-numbers (since each such will render it true, for X includes ℕ in its range), but (i) will. Hence, under this assumption, (ii) is superfluous and (i) will do by itself. On the other hand, if the range of the second-order quantifier is taken to consist only of finite sets of all sizes, then (i) will not suffice to exclude all non-numbers (since the antecedent of its instances will be false, if k is not a natural number), whereas (ii) will. Hence, in this second situation, (ii) by itself suffices and (i) is not needed. Offering both clauses, as Wang does, will inevitably be redundant, but may be appropriate if one wishes to provide a definition that works whether or not infinite sets are included in the range of the second-order quantifier.

We are now in a position to resolve the conflict presented in the first paragraph of this note. If one is imagining that second-order quantifiers range over infinite sets, then Dummett's definition effectively consists in clause (i) and Wang is correct to assimilate it to Frege–Dedekind's PD definition: for the elimination of non-natural numbers will require ℕ itself to be in the domain of second-order quantification. On the other hand, if one takes these quantifiers to range only over all finite sets, as Feferman and Hellman do, then Dummett's definition in effect amounts to (ii) and therefore, for reasons given in the previous note, should be likened rather to the BU definition.

In this context, it is worth mentioning another definition of the natural numbers, this time first offered by Quine (1961); see also (Quine 1969, pp. 75ff.). According to this definition, $k \in \mathbb{N}$ just in case

(iii) $(\forall X)((k \in X \,\&\, (\forall y)(S(y) \in X \rightarrow y \in X)) \rightarrow 0 \in X)$.

If we assume that the second-order quantifier ranges over infinite sets, then this definition captures all the natural numbers (since if k is a natural number then every set containing k and closed under predecessor will contain 0) and only them (for if k is not a natural number, say an entity with infinitely many predecessors, then the complement of $\mathbb{N}$ contains k and is closed under predecessor, but does not contain 0, and hence the closure of $\{k\}$ under predecessor will not contain 0).

However, contrary to Quine's suggestion, the definition does not work to exclude non-numbers if the range of the second-order quantifier is restricted to finite sets of all sizes: for if k has infinitely many predecessors, then (iii) will be vacuously true. If the domain is so restricted, then (iii) could be supplemented by (iv):

(iv) $(\exists X)(k \in X \;\&\; (\forall y)(S(y) \in X \rightarrow y \in X))$.

It is certainly true that if k is a natural number, then there exists a finite set containing k and closed under predecessor. Furthermore, if k has infinitely many predecessors, then it fails to satisfy (iv), for there will not exist a finite set of the requisite kind. Can one, in this context, make do with (iv) alone? No, for (iv) fails to rule out Caesar as a natural number, because there does exist a finite set containing Caesar and closed under predecessor, namely {Caesar}. But, taking X to be this set, we see that Caesar does not satisfy (iii). In sum, if the second-order quantifier ranges only over all finite sets, then one emendation of Quine's definition consists of the conjunction of (iii) and (iv). (See (George 1987) and (Parsons 1987, pp. 210–1).)

It is not easy to say where this particular emendation of Quine's definition falls in our classificatory scheme, for it contains elements of both the PD and the BU approaches. There is, however, another way of supplementing (iii) that leads to a more straightforward outcome. Consider:

(iv′) $(\exists X)(k \in X \;\&\; (\forall y)(S(y) \in X \rightarrow y \in X) \;\&$
$(\forall y)((y \in X \;\&\; y \neq 0) \rightarrow (\exists z)(S(z) = y)))$.

Clearly, if k is a natural number, it satisfies (iv′). Also, (iv′) rules out all non-natural numbers, including Caesar, since he is unequal to 0 and has no predecessor. Hence, when the second-order quantifier ranges only over all finite sets, (iii) is superfluous and can be replaced by (iv′). Furthermore, (iv′) is plainly in the spirit of a BU approach to the natural numbers.

Although Feferman and Hellman say that their own approach realizes Dummett's definition, it is in fact closer to (iv′) than it is to (ii), for their definition employs closure under predecessor rather than closure under successor. (Their definition differs from (iv′) in only one respect: they require that the set X be the *smallest* set containing k and closed under predecessor. However, an examination of the proof of their Theorem 7 shows that this additional requirement plays no role in the proof, and therefore could have been eliminated. Thus, (iv′) captures the essence of Feferman and Hellman's definition.)

Recently, Peter Aczel has shown that the existence of a structure satisfying Peano's axioms can be proven in Feferman and Hellman's system *without* using their axiom (Card), which restricts the range of the second-order quantifiers to

Dedekind finite sets (Feferman, personal communication). Aczel's approach is, roughly, to let the universe of the structure be defined by the conjunction of (iii) and (iv), rather than (iv′). This conjunction characterizes the natural numbers whether or not the range of the second-order quantifiers includes infinite sets. However, if infinite sets are included then, as we observed earlier in this note, the exclusion of non-numbers by (iii) requires reference to a collection, the complement of $\mathbb{N}$, that is defined in terms of $\mathbb{N}$, the very set being defined. Thus, despite Feferman and Hellman's description of their formal system as "predicatively justified," it seems questionable to us whether Aczel's definition should be called predicative. Aczel's theorem can also be proven using the conjunction of (i) and (ii), rather than the conjunction of (iii) and (iv). However, if infinite sets are not excluded from the range of the second-order quantifiers, then the use of (i) will once again render the definition impredicative.

11. There is a circularity in this argument, for mathematical induction will be needed in order to show that P holds at every stage of the construction. This is not a circularity in the BU definition of natural number; rather, it is a circularity that appears in the justification of induction on the basis of that definition. (It is reminiscent of the circularity Hume discovered in attempting to justify empirical induction.) Yet the argument is, as Parsons has noted, "no worse than arguments for the validity of elementary logical rules" (1992, p. 143).

One way of summarizing both this argument for induction from the BU definition and the argument for why the PD definition captures only the natural numbers (see above) is to say that the extremal clause and the principle of mathematical induction are interderivable. For example, this is essentially what S. C. Kleene says (1952, p. 22). While correct as far as it goes, we prefer not to put the matter this way, for it obscures the distinctive approaches to the natural numbers that we believe animate the two definitions.

Even more obscuring is to construe the extremal clause as *saying* that induction is valid. Parsons, for example, at one time articulated the view that the principle of induction "could be regarded simply as an interpretation of" the extremal clause. Yet, he did not then advance the position and in fact also mentioned the possibility that "the induction principle [...] will be in some way a consequence of" the extremal clause (Parsons 1967, p. 194). More recently, however, he appears to endorse this view, as when he describes induction as "a principle cashing in our intention that the numbers should be what is obtained by the introduction rules and those alone" (Parsons 1992, p. 143). This way of viewing the matter no doubt contributes to his analysis of possible options:

> The readily available alternatives to something like the induction-definition model of the concept of natural number [...] would be to give it an explanation that is blatantly circular, such as, the natural numbers are what is obtained by beginning with 0 and iterating the successor operation an *arbitrary finite number* of times, or to take the concept of natural number as given and the principle of induction as evident without any explication connecting it with the concept of

> natural number. Either alternative seems to me a counsel of philosophical despair that leaves us with no motivation for the principle of induction. (Parsons 1992, p. 143)

Yet, as we have seen, these alternatives are not exhaustive, for the proponents of the BU definition view it as neither "the induction-definition model of the concept of natural number" (i.e., the PD definition), nor as circular, nor as failing to provide a justification of induction.

Again, we are not arguing in favor of one or the other definition, but rather attempting to delineate two philosophically distinct approaches to the nature of number.

12. For example, this thought may be behind Parsons' claim that "If one explains the notion of natural number in such a way that induction falls out of the explanation, then one will be left with a similar impredicativity"—similar, that is, to the impredicativity of the PD definition (Parsons 1992, p. 141). For a discussion, see (George 1987).

13. Parsons gives voice to this view as well when he suggests that "induction is constitutive of the meaning of the term 'natural number'" (Parsons 1992, p. 155). Dummett, also, believes that "the meaning of the expression 'natural number' involve[s], not only the criterion for recognising a term as standing for a natural number, but also the criterion for asserting something about all natural numbers" (Dummett 1978, p. 194).

14. Edward Nelson, for example, adopts a principle of induction restricted to those predicates involving bounded quantification (Nelson 1986, p. 2).

Another example is (Feferman and Hellman 1995), in which the authors establish induction only for formulas containing no quantification over the collection of all classes. In fact, it can be shown that induction for all formulas is not provable in their system EFSC*. The reason is that, if full induction were added to EFSC* as a new axiom, it would be possible to define a satisfaction relation and use it to prove that all theorems of PA are true in $\mathbb{N}$, and therefore that PA is consistent. But as Feferman and Hellman show in their Metatheorem 9, the consistency of PA is not provable in EFSC*.

It might be helpful to spell out a few more of the details of this proof. The satisfaction relation for formulas in the language of PA can be represented as a function assigning to each pair (φ, s), where φ is a formula and s is an assignment of values to the free variables of φ, one of the values 1 or 0, representing true and false. By assigning Gödel numbers to both formulas and assignments of values to variables, we can think of this function as mapping $\mathbb{N} \times \mathbb{N}$ to $\{0, 1\}$. Let us say that a function from $\{0, 1, 2, \dots, n\} \times \mathbb{N}$ to $\{0, 1\}$ is an *n-satisfaction function* if it satisfies the usual recursive definition of satisfaction for formulas with Gödel numbers up to n. Then we can prove by induction that $(\forall n \in \mathbb{N})(\exists F)(F$ is an n-satisfaction function). Note that the formula being proven by induction contains the quantifier "$(\exists F)$", where "F" stands for an infinite class, so the weak form of induction proven by Feferman and Hellman would not be sufficient for this proof. Our satisfaction relation $\mathrm{Sat}(x, y)$ can now be defined to be the formula: $(\exists F)(F$ is an x-satisfaction function & $F(x, y) = 1)$.

It is interesting that full induction is provable in a slight strengthening of Feferman and Hellman's system. Let EFSC^+ be the same as EFSC^*, except with no restrictions on the formula φ in (Sep), the separation axiom for finite sets. Then we can prove full induction in EFSC^+ by imitating the proof of Feferman and Hellman's Theorem 8. Let $\varphi(n)$ be any formula, and suppose we have both $\varphi(0)$ and $(\forall n)(\varphi(n) \rightarrow \varphi(n+1))$. If $\neg\varphi(m)$ for some natural number m, then let $B = \{ n : n \leq m \}$, a finite set since m is a natural number, and let $Y = \{ n \in B : \neg\varphi(n) \}$. Then, as in Feferman and Hellman's proof of their Theorem 8, it can be shown that Y is both finite and Dedekind-infinite, contradicting the cardinality axiom of EFSC^+. Note that the definition of Y requires the strengthened version of (Sep), since φ might involve quantification over the collection of all classes.

It is difficult to say whether or not EFSC^+ should be considered predicative. The strengthened version of (Sep) allows one to quantify over the collection of all classes when defining a subset of a finite set, and such a definition would appear to be impredicative. But even the original version of (Sep) allows one to define a subset of a finite set by quantifying over the collection of all finite sets, a collection that includes the very set being defined, and Feferman and Hellman do not consider this to be impredicative. Their reason is that they "assume that the notion of *finite set* is predicatively understood, governed by some elementary closure conditions" (p. 2). Feferman and Hellman appear to take this to mean that a definition of a set that involves quantification over the collection of all finite sets is predicative, but one involving quantification over the collection of all classes is not. Thus, they consider their version of (Sep) to be predicatively acceptable, but would presumably reject the strengthened version of (Sep) as impredicative.

Some might question whether such closure conditions for the collection of finite sets should be considered to be predicatively acceptable. But if closure conditions are to be accepted, we believe an argument can be made for a closure condition that implies the strengthened version of (Sep). One way of justifying such a closure condition would be to argue that given a finite set, it is possible to make a finite list of all subsets of that set. Each of these subsets is definable by a predicative definition that simply lists its elements. Any other definition of a subset of the original finite set, even an impredicative one, must pick out one of these subsets, whose existence has already been established by a predicative definition. Thus, for any definition that specifies unambiguously which elements of the finite set are to be included in a subset, there must exist a subset containing precisely the elements specified by that definition. This reasoning would apply whether the definition involves quantification over only the collection of all finite sets or quantification over the collection of all classes, so it would justify not only the original version of (Sep), but also the strengthened version.

15. For example, see the quotation from Parsons in note 12; see also (Nelson

1986, pp. 1-2).

16. This term is suggested by Ludwig Wittgenstein's (1921, 3.263).

17. Thus, we dissent from what seems to be Michael Dummett's position in (Dummett 1993*b*); see p. 443. For some further discussion, see (George 1994).

Bibliography

Dedekind, R. (1967). Letter to Keferstein. In *From Frege to Gödel: a source book in mathematical logic, 1879–1931* (ed. J. van Heijenoort), pp. 98–103. Harvard University Press, Cambridge, Massachusetts.

Dummett, M. A. E. (1978). The philosophical significance of Gödel's theorem. Reprinted in his *Truth and other enigmas*, pp. 186–201. Harvard University Press, Cambridge, Massachusetts.

Dummett, M. A. E. (1993*b*). What is mathematics about? Reprinted in his *The seas of language*, pp. 429–45. Clarendon Press, Oxford

Enderton, H. (1972). *A mathematical introduction to logic.* Academic Press, New York.

Feferman, S. and Hellman G. (1995). Predicative foundations of arithmetic. *Journal of Philosophical Logic*, **24**, 1–17.

George, A. (1987). The imprecision of impredicativity. *Mind*, **96**, 514–18.

George, A. (1994). Intuitionism and the poverty of the inference argument. *Topoi*, **13**, 79–82.

Isaacson, D. (1987). Arithmetical truth and hidden higher-order concepts. In *Logic Colloquium '85* (ed. Paris Logic Group), pp. 147–69. North-Holland, Amsterdam.

Kleene, S. C. (1952). *An introduction to metamathematics.* North-Holland, Amsterdam and von Nostrand, Princeton.

Nelson, E. (1986). *Predicative arithmetic.* Mathematical Notes, Vol. 32, Princeton University Press.

Parsons, C. (1967). Mathematics, foundations of. In *The encyclopedia of philosophy*, Vol. 5 (ed. P. Edwards), pp. 188–213. Macmillan, New York.

Parsons, C. (1987). Developing arithmetic in set theory without infinity: some historical remarks. *History and Philosophy of Logic*, **8**, 201–13.

Parsons, C, (1992). The impredicativity of induction. In *Proof, logic and formalization* (ed. M. Detlefsen), pp. 139–61. Routledge, London.

Poincaré, H. (1952). *Science and method* (trans. F. Maitland). Dover, New York.

Quine, W. V. (1961). A basis for number theory in finite classes. *Bull. American Math. Soc.*, **67**, 391–92.

Quine, W. V. (1969). *Set theory and its logic.* Harvard University Press, Cambridge, Massachusetts.

Wang, H. (1963). Eighty years of foundational studies. Reprinted in his *A survey of mathematical logic*, pp. 34–56. North-Holland, Amsterdam.

Wittgenstein, L. (1981). *Tractatus logico-philosophicus.* Routledge, London.

Department of Philosophy
Amherst College
Amherst, MA 01002
USA
email: AGEORGE@amherst.edu

Department of Mathematics
and Computer Science
Amherst College
Amherst, MA 01002
USA
email: DJVELLEMAN@amherst.edu

18
The tower of Hanoi

W. Hugh Woodin

1 Introduction

A concern perhaps more prevalent among logicians than among mathematicians in general is that the formal axioms chosen for set theory lead to a contradiction. Further, possible inconsistency has been a frequent point in the criticism of the study of large cardinals in general. This is without question a valid concern; however the intricate structure that has evolved in the course of this study seems to provide justification for the belief in the consistency of these axioms.

By Gödel's Second Incompleteness Theorem, any system of axioms of *reasonable* expressive power is subject to the possibility of inconsistency, including of course the axioms for number theory. A natural question is how profound an effect could an inconsistency have on our view of mathematics or indeed on our view of physics.

Define an iterated exponential function $\mathrm{Exp}_k(n)$ as follows by induction on k:

(1) $\mathrm{Exp}_1(n) = 2^n$;

(2) $\mathrm{Exp}_{k+1}(n) = 2^{\mathrm{Exp}_k(n)}$.

By a routine Gödel sentence construction we shall produce a formula $\Omega(x_1)$ in the language for set theory which implicitly defines a property for finite sequences. For a given sequence this property is easily decided; if s is a sequence of length n with this property, then s is a sequence of non-negative integers each less than n and the verification can be completed (with appropriate inputs) in significantly fewer than n^2 steps.

If there exists a sequence of length n with this property, then

$$\mathrm{Exp}_{2011}(n)$$

does *not* exist, in the sense that there is a finite state Turing machine which cannot halt (the machine has no final state), and yet cannot run for this many steps. This machine is easily specified in advance and does not depend on n.

The philosophical consequences of the existence of such a sequence are clearly profound, for it demonstrates the necessity of the finiteness of the universe. Clearly such a sequence should not exist. However this sentence has the feature that, if abitrarily large sets *can* exist, then, for each suitable n, there is no proof of *length* less than n that no such sequence of length n can have this property.

We shall make these claims more precise.

Can such finite sequences exist? In the universe of sets there are very reasonable worlds (that is, models) in which they do exist, though in these models the sequences have nonstandard length (of course). Perhaps our experience in mathematics to date refutes the existence of such sequences with more accessible length. It would be comforting to know that no one will by any means find such a sequence of length say 10^{24}. A consequence of quantum mechanics (as opposed to classical mechanics) is that one could build a device with a *non-zero* (though ridiculously small) chance of finding such a sequence, if such a sequence exists.

We shall argue that there are limitations to the extent our experience in mathematics to date refutes the existence of such sequences. In fact we shall argue that a consistent philosophical view must in effect acknowledge the possiblity that the sequences of length 10^{24} could exist, just as those who study *large cardinals* must admit the possibility that the notions are not consistent.

There are related mathematical questions which we shall also discuss.

2 Preliminaries

There are some metamathematical considerations to be dealt with. We shall work as mathematicians. We assume that the universe of sets exists and that the accepted axioms hold for this universe. Our constructions refer to objects in this universe. This is simply a device to facilitate our discussions. Our final product is a concrete object, specifically a formula which we could for amusement explicitly write down.

We detour and develop the basics of mathematical logic that we shall need. We shall assume familiarity with set theory at a naïve level, though we recall the basic axioms, as follows.

Axiom 0 There exists a set.

Axiom 1 (Extensionality) Two sets A and B are equal if and only if they have the same elements.

Axiom 2 (Pairing) If A and B are sets, then there exists a set $C = \{A, B\}$ whose only elements are A and B.

Axiom 3 (Union) If A is a set, then there exists a set C whose elements are the elements of the elements of A.

Axiom 4 (Power set) If A is a set, then there exists a set C whose elements are the subsets of A.

Axiom 5 (Regularity or Foundation) If A is a set, then either A is empty (i.e., A has no elements) or there exists an element C of A which is disjoint from A.

Axiom 6 (Comprehension) If A is a set and $P(x)$ formalizes a property of sets, then there exists a set C whose elements are the elements of A with this property.

Axiom 7 (Replacement) If A is a set and $P(x, y)$ formalizes a property of pairs of sets which defines a function of sets, then there exists a set C which contains as elements all the values of this function acting on the elements of A.

Axiom 8 (Axiom of Choice) If A is a set whose elements are pairwise disjoint and each nonempty, then there exists a set C which contains exactly one element from each element of A.

Axiom 9 (Infinity) There exists a set W which is nonempty and such that, for each element A of W, there exists an element B of W such that A is an element of B.

We make some remarks.

Axiom 6 and **Axiom 7** are really infinite lists or schemata corresponding to the possibilities of the *acceptable properties*. These axioms are vague in that it may not be clear what an acceptable property is. Intuitively these properties are those that can be expressed using only the fundamental relationships of equality and set membership. This becomes clear with an understanding of elementary mathematical logic.

Axioms 0–8 are (essentially) a reformulation of the axioms of number theory. It is the **Axiom of Infinity** that takes one from number theory to set theory. An exact reformulation of number theory is given by **Axioms 0–8** together with the negation of **Axiom 9** and a strengthened form of **Axiom 5** (which is necessary to eliminate certain sets).

Mathematical constructions specify objects in the universe of sets; this is our informal point of view. For example, by using a property that cannot be true for any set, namely $x \neq x$, one can easily show using **Axiom 0** and **Axiom 4** that there exists a set with no elements. By **Axiom 1** this set is unique; it is the *empty set* and is denoted by $\emptyset$.

We discuss several examples of how familiar mathematical notions are formalized as sets. If A and B are sets, then the ordered pair (A, B) is the set $\{\{A\}, \{A, B\}\}$. This set exists by three applications of **Axiom 2**. This definition of an ordered pair is due to von Neumann, and it is easily verified that this definition performs as desired; i.e., if $(A, B) = (C, D)$, then $A = C$ and $B = D$. Now it is routine to define the notion of a function as a set of ordered pairs, with the obvious constraints.

We define those sets which we view to be the non-negative integers or whole numbers. A set A is *transitive* if each element of A is also a subset of A. A set A is an *ordinal* if it is transitive and has the additional property that, if B, C are elements of A, then $B \in C$, $B = C$, or $C \in B$. (This definition originates with von Neumann.) It is not difficult to show that, if A and B are ordinals, then $A \in B$, $A = B$, or $B \in A$. For this, one uses **Axiom 5** and **Axiom 6**. Thus the relationship of membership linearly orders the ordinals. If A and B are ordinals we write $A < B$ to indicate that $A \in B$. Suppose that A is an ordinal. Then

$$A = \{B \mid B \text{ is an ordinal and } B < A\};$$

each ordinal is simply the set of ordinals which precede it in the given order. For each ordinal A, there exists an ordinal B such that $A < B$ and such that, if C is an ordinal with $A < C$, then either $B = C$ or $B < C$. The ordinal B is the *successor* of A and it is denoted by $A + 1$. It is easily verified that

$$A + 1 = A \cup \{A\}.$$

Thus the natural definition of a *finite ordinal* is as follows. An ordinal A is finite if, for each $B \in A$, either

$$B = \emptyset \quad \text{or} \quad B = C + 1 \ \text{ for some } \ C \in A.$$

Clearly the finite ordinals form an initial segment of the ordinals. The first four ordinals are:

$$0, \{0\}, \{0, \{0\}\}, \{0, \{0\}, \{0, \{0\}\}\}.$$

These are the numbers $0, 1, 2, 3$. More generally the finite ordinals are the non-negative integers. The **Axiom of Infinity** in conjunction with the other axioms implies the existence of an infinite ordinal. The least infinite ordinal is denoted by ω. The finite ordinals are exactly the ordinals A such that $A < \omega$.

We now define finite sequences. Suppose that M is a set and that n is a non-negative integer (i.e., a finite ordinal). The set of n-tuples of elements of M is the set M^n of all functions

$$f : n \longrightarrow M.$$

The finite sequences of elements of M is the set

$$M^{<\omega} = \bigcup \{M^n \mid n \in \omega\}.$$

If s and t are finite sequences of elements of M, then $s + t$ is the finite sequence defined by concatenating s and t. Thus, if $s \in M^i$ and if $t \in M^j$, then we define $s + t \in M^{i+j}$ by

$$(s + t)(k) = \begin{cases} s(k) & \text{if } k < i, \\ t(k - i) & \text{if } k \geq i. \end{cases}$$

3 First-order logic

We give a brief development of formal logic. In this we continue to discuss objects in the universe of sets. Formal logic involves the definition of *language* and the definition of *proof* for the language. Intuitively the language consists of certain *expressions* involving symbols from an *alphabet*. Formally, the alphabet is the set of finite ordinals, the first seven of which are regarded as the logical symbols and as such are denoted by

$$\hat{\in} \ \ \hat{=} \ \ (\ \ ) \ \ \vee \ \ \exists \ \ \neg,$$

and the remaining are the *variable symbols* and as such are denoted in increasing order by

$$x_0, x_1, x_2, \ldots, x_k, \ldots .$$

The expressions of our language are certain finite sequences of elements of the alphabet. We adopt the usual conventions and write for example

$$(x_1 \hat{=} x_1)$$

to indicate the 5-tuple

$$\langle (, x_1, \hat{=}, x_1,) \rangle ,$$

which is really the function

$$f : 5 \rightarrow \omega$$

given by $f(0) = 2$, $f(1) = 8$, $f(2) = 1$, $f(3) = 8$, and $f(4) = 3$.

Also, if ψ and φ are finite sequences of elements of the alphabet, then

$$(\psi \vee \varphi)$$

indicates the finite sequence

$$\langle (\rangle + \langle \psi \rangle + \langle \vee \rangle + \langle \varphi \rangle + \langle) \rangle .$$

The expressions of the language are *formulas*; these are defined by induction. We begin with the *atomic formulas.* These are the 5-tuples of the form

$$(x_i \hat{=} x_k)$$

or of the form

$$(x_i \hat{\in} x_k).$$

The formulas are generated from the atomic formulas by the following operations or rules of formation.

Connective Rule Suppose that ψ and φ are formulas. Then

$$(\psi \vee \varphi) \quad \text{and} \quad (\neg \psi)$$

are formulae.

Quantifier Rule Suppose that ψ is a formula and that x_i is a variable. Then

$$(\exists x_i \psi)$$

is a formula.

Suppose that ψ and φ are formulas. The formula ψ is a *subformula* of φ if it is a *consecutive* subsequence of φ. For example consider the formula

$$((x_1 \hat{=} x_2) \vee (\exists x_2 (x_2 \hat{=} x_3))).$$

Thus $(x_1 \hat{=} x_2)$ is a subformula of this formula, as is the formula $(x_2 \hat{=} x_3)$. Note that the formula $(x_1\hat{=}x_3)$ is not a subformula, though it is a subsequence of the given formula.

It is not difficult to show that the subformulas of a formula ψ are exactly the formulas involved in the formation of ψ. Thus the formula ψ has a unique presentation as a finite sequence generated from the atomic formulas using the rules of formation.

Suppose that ψ is a formula. An occurrence of $\exists x_i$ in ψ is an occurrence of $\exists$ immediately followed by an occurrence of x_i. The *scope* of an occurrence of $\exists x_i$ in ψ is the unique subformula defined by that occurrence of $\exists x_i$. For example in the formula

$$((x_1\hat{=}x_2) \vee (\exists x_2(x_2\hat{=}x_3))),$$

the first occurrence of x_2 is not in the scope of any occurrence of $\exists x_2$. The second occurrence of x_2, as well as the first occurrence of x_3, are each within the scope of the first occurrence of $\exists x_2$.

All of this serves to motivate the following definition. An occurrence of a variable x_i in a formula ψ is *free* if it is not within the scope of any occurrence of $\exists x_i$; otherwise the occurrence of x_i is *bound.* A variable x_i is a *free variable* of ψ if there is a free occurrence of x_i in ψ.

A formula ψ is a *sentence* if ψ has no free variables.

We shall write $\psi(x_0, \ldots, x_n)$ to indicate that ψ is a formula such that every free variable of ψ is in the set $\{x_0, x_1, \ldots, x_n\}$.

4 A definition of truth

A *structure* or *model* for our language is a pair

$$\mathcal{M} = (M, E)$$

such that M is a *non-empty* set and such that E is a binary relation on M; i.e.,

$$E \subset M \times M,$$

where $M \times M$ is the set of ordered pairs of elements of M.

Every set naturally defines a model for our language by defining E to be the binary relation of set membership restricted to the set. For example, if X is a non-empty set then the model corresponding to X is the pair

$$(X, E),$$

where

$$E = \{(a, b) \mid a \in X,\ b \in X, \text{ and } a \in b\}.$$

We denote the model (X, E) by $(X, \in)$. Models of the form $(X, \in)$ are *standard models.*

Suppose that

$$\mathcal{M} = (M, E)$$

is a model, ψ is a formula with its free variables contained in the set

$$\{x_0, x_1, \dots, x_n\} ,$$

and that $a_0, a_1, \dots, a_n$ are elements of M. There is a natural interpretation of the truth of ψ in $\mathcal{M}$ at $\langle a_0, \dots, a_n\rangle$. This arises from interpreting the symbol '$\widehat{=}$' by the relation of equality, the symbol '$\widehat{\in}$' by the relation E, '$\vee$' by 'or', '$\neg$' by 'not', '$\exists$' by 'there exists', and for each $i \le n$, 'x_i' by a_i at each free occurrence of x_i in ψ.

We write

$$\mathcal{M} \models \psi[a_0, \dots, a_n]$$

to indicate that ψ is true in $\mathcal{M}$ at $\langle a_0, \dots, a_n\rangle$.

If ψ is a sentence, then the truth of ψ in $\mathcal{M}$ does not depend on the choice of $\langle a_0, \dots, a_n\rangle$, and we can unambiguously define ψ to be true in $\mathcal{M}$ or false in $\mathcal{M}$. In this case we write

$$\mathcal{M} \models \psi$$

to indicate that ψ is true in $\mathcal{M}$.

We give a more precise definition of

$$\mathcal{M} \models \psi[a_0, \dots, a_n]$$

as follows.

Suppose that S is a set of formulas and that S is closed under subformulas; i.e., if $\psi \in S$ and if φ is a subformula of ψ, then $\varphi \in S$. Suppose that n is such that every variable occurring in some formula of S is in the set $\{x_0, \dots, x_n\}$. Further, suppose that $\mathcal{M} = (M, E)$ is a model. An *oracle* for the pair $(\mathcal{M}, S)$ is a function

$$I : S \times M^{n+1} \to \{0, 1\}$$

such that, for all $\langle a_0, \dots a_n\rangle \in M^{n+1}$, the following hold.

Atomic Case (1) If $(x_i \widehat{=} x_j) \in S$, then

$$I((x_i \widehat{=} x_j), \langle a_0, \dots, a_n\rangle) = 1 \text{ if and only if } a_i = a_j.$$

(2) If $(x_i \widehat{\in} x_j) \in S$, then

$$I((x_i \widehat{\in} x_j), \langle a_0, \dots, a_n\rangle) = 1 \text{ if and only if } (a_i, a_j) \in E.$$

Connective Case (1) If $(\psi_1 \vee \psi_2) \in S$, then

$$I((\psi_1 \vee \psi_2), \langle a_0, \dots, a_n\rangle) = 1$$

if and only if

$$I(\psi_1, \langle a_0, \dots, a_n\rangle) = 1 \text{ or } I(\psi_2, \langle a_0, \dots, a_n\rangle) = 1.$$

(2) If $(\neg\psi) \in S$, then

$$I((\neg\psi), \langle a_0, \dots, a_n\rangle) = 1 \text{ if and only if } I(\psi, \langle a_0, \dots, a_n\rangle) = 0.$$

Quantifier Case If $(\exists x_i \psi) \in S$, then

$$I((\exists x_i \psi), \langle a_0, \dots, a_n\rangle) = 1$$

if and only if

$$I(\psi, \langle a_0, \dots, a_{i-1}, a, a_{i+1}, \dots a_n\rangle) = 1$$

for some $a \in M$.

It is routine to check that the function I exists and is uniquely specified. Further, if $I_1 : S_1 \times M^{n+1} \to \{0,1\}$ and $I_2 : S_2 \times M^{n+1} \to \{0,1\}$ are oracles for $(\mathcal{M}, S_1)$ and $(\mathcal{M}, S_2)$, respectively, and if $S_1 \subset S_2$, then I_1 is the restriction of I_2 to $S_1 \times \mathcal{M}^{n+1}$. Similarly, if $I_1 : S \times M^{n_1+1} \to \{0,1\}$ and $I_2 : S \times M^{n_2+1} \to \{0,1\}$ are oracles for $(\mathcal{M}, S)$, and if $n_1 \leq n_2$, then

$$I_1(\psi, \langle a_0, \dots, a_{n_1}\rangle) = I_2(\psi, \langle a_0, \dots, a_{n_2}\rangle)$$

for all $\psi \in S$ and $\langle a_0, \dots, a_{n_2}\rangle \in M^{n_2+1}$.

Suppose that ψ is a formula, and let S be the set of subformulas of ψ (including ψ). Suppose also that $I : S \times M^{n+1} \to \{0,1\}$ is an oracle for $(\mathcal{M}, S)$ and that $\langle a_0, \dots, a_n\rangle \in M^{n+1}$. Then we define

$$\mathcal{M} \models \psi[a_0, \dots, a_n]$$

if $I(\psi, \langle a_0, \dots, a_n\rangle) = 1$. This is well-defined, and the relation

$$\mathcal{M} \models \psi[a_0, \dots, a_n]$$

depends only on those a_i for which the variable x_i is a free variable of ψ. Thus, if ψ is a sentence, then

$$\mathcal{M} \models \psi[a_0, \dots, a_n]$$

does not depend on $\langle a_0, \dots, a_n\rangle$ and is independent of n. We leave the verification of these claims to the dedicated reader. Our purpose here was the definition of an oracle, a notion we shall require later.

We give some trivial examples. Suppose first that ψ is the sentence

$$(\exists x_1 (\exists x_2 (\neg (x_1 \mathrel{\widehat{=}} x_2))))$$

and that $\mathcal{M} = (M, E)$ is a model. Then $\mathcal{M} \models \psi$ if and only if M contains at least two distinct elements. Secondly, the sentence

$$(\exists x_1 (\neg (x_1 \mathrel{\widehat{=}} x_1)))$$

is not true in any model.

5 A definition of proof

We denote by $\mathcal{L}(\widehat{=}, \widehat{\in})$ the set of all formulas of our language. We remark that this set *exists* as a consequence of our axioms, specifically the **Axiom of Infinity**.

A *theory* is a subset of $\mathcal{L}(\widehat{=}, \widehat{\in})$ which contains only sentences.

Suppose that T is a theory and that ψ is a sentence. We define the notion that the theory T *proves* ψ. We write

$$T \vdash \psi$$

to indicate that T proves ψ.

It is useful to introduce some standard abbreviations. The first expresses 'implies' in terms of 'not' and 'or', the second expresses 'and' in terms of 'not' and 'or' and the third expresses 'if and only if' in terms of 'not' and 'or'.

Suppose that ψ and φ are formulas. Then

$$\psi \rightarrow \varphi$$

indicates the formula $((\neg\psi) \vee \varphi)$,

$$(\psi \wedge \varphi)$$

indicates the formula $(\neg((\neg\psi) \vee (\neg\varphi)))$, and, using these abbreviations,

$$(\psi \leftrightarrow \varphi)$$

indicates the formula

$$((\psi \rightarrow \varphi) \wedge (\varphi \rightarrow \psi)).$$

Suppose that ψ is a formula and x_i is a variable. We express 'for all' in terms of 'there exists' and 'not', letting

$$(\forall x_i \psi)$$

indicate the formula $(\neg(\exists x_i(\neg\psi)))$.

Suppose that S is a set of formulas. A formula ψ is obtained from S by *modus ponens* if there exists a formula $\varphi \in S$ such that $(\varphi \rightarrow \psi) \in S$.

Suppose that T is a theory. A *proof* from T is a finite sequence

$$\langle \psi_0, \ldots, \psi_n \rangle$$

of formulas such that, for all $i \leq n$, $\psi_i \in T$, ψ_i is a *logical axiom*, or ψ_i is obtained from $\{\psi_k \mid k < i\}$ by modus ponens.

Note that, if $\langle \psi_0, \ldots, \psi_n \rangle$ is a proof from T, then ψ_0 and ψ_1 are necessarily elements of T or logical axioms.

It remains to specify the logical axioms. To avoid a technical digression we defer this to the appendix. Intuitively the logical axioms are formulas $\psi(x_0, \ldots, x_k)$ such that for trivial reasons $\mathcal{M} \models \psi[a_0, \ldots, a_k]$ for any model $\mathcal{M} = (M, E)$ and for any $\langle a_0, \ldots, a_k \rangle \in M^{k+1}$. For example, the formula $(x_1 \hat{=} x_1)$ is a logical axiom. We restrict the list slightly so that the verification that a formula of length n is a logical axiom is relatively simple, as is the custom in computer science.

Suppose that T is a theory and that ψ is a sentence. Then

$$T \vdash \psi$$

if there is a proof $\langle \psi_0, \ldots, \psi_n \rangle$ from T such that $\psi_n = \psi$.

Suppose that T is a theory and that $\mathcal{M}$ is a model. Then we write $\mathcal{M} \models T$ to indicate that $\mathcal{M} \models \varphi$ for all $\varphi \in T$. Suppose that ψ is a sentence. We write

$$T \models \psi$$

to indicate that, for all models $\mathcal{M}$, $\mathcal{M} \models \psi$ whenever $\mathcal{M} \models T$.

The connection between the notions of 'proof' and 'truth' is the relation between $T \vdash \psi$ and $T \models \psi$.

Theorem 5.1 (Gödel's completeness theorem) *Suppose that $T \subset \mathcal{L}(\hat{=}, \hat{\in})$ is a theory and that $\psi \in \mathcal{L}(\hat{=}, \hat{\in})$ is a sentence. Then the following are equivalent:*

(1) $T \vdash \psi$;

(2) $T \models \psi$. □

6 The formula

Our pathological sequences will be associated to finite integers n of the form 10^{24k}. These will be specified by producing for each k a sentence of our language, $\mathcal{L}(\hat{=}, \hat{\in})$. For technical reasons it is important that the sentence we indirectly associate to n be 'short'. These constructions require the following technical devices.

The Σ_0-*formulas* are the formulas generated from the atomic formulas using the operations given by the **Connective Rule** and the following modified version of the **Quantifier Rule**.

Bounded Quantifier Rule Suppose that ψ is a formula, x_i is a variable, and that x_j is a variable. Then

$$(\exists x_i((x_i \hat{\in} x_j) \wedge \psi))$$

is a formula.

Suppose that ψ is a formula. The *universal closure* of ψ is the sentence ψ^*, obtained from ψ as follows. Let $x_{n_1}, \ldots, x_{n_k}$ be the free variables of ψ. Then ψ^* is the formula indicated by

$$(\forall x_{n_1}(\ldots(\forall x_{n_k}\psi)\ldots)).$$

Suppose that ψ is a formula and that x_i is a variable with no occurrences in ψ. Let $\psi[x_i{:}x_j]$ be the formula obtained from ψ by substituting x_i for each free occurrence of x_j. For example, if ψ is the formula

$$(\exists x_0(x_1 \hat{=} x_0)) \vee (x_0 \hat{\in} x_2)),$$

then $\psi[x_3{:}x_0]$ is the formula

$$((\exists x_0(x_1 \hat{=} x_0)) \vee (x_3 \hat{\in} x_2)).$$

We first define the theory with which we shall be working; it is our base theory, and we denote it by T_0. The elements of T_0 are the universal closures of the formulas generated as follows.

Axiom I (Extensionality):

$$(\forall x_1(\forall x_2((x_1\hat{=}x_2) \leftrightarrow (\forall x_3((x_3\hat{\in}x_1) \leftrightarrow (x_3\hat{\in}x_2))))))$$

Axiom II (Σ_0-Regularity):

$$(\exists x_2((\exists x_0((x_0\hat{\in}x_1)\wedge\psi))\rightarrow(((x_2\hat{\in}x_1)\wedge\psi_2)\wedge(\forall x_4(\neg((x_4\hat{\in}x_1)\wedge(\psi_4\wedge(x_4\hat{\in}x_2))))))$$

where ψ is a Σ_0-formula with no occurrences of the variables x_2 and x_4, ψ_2 is the formula $\psi[x_2{:}x_0]$, and ψ_4 is the formula $\psi[x_4{:}x_0]$.

Axiom I is the **Axiom of Extensionality** and is, as given, a sentence. The axioms generated by **Axiom II** are forms of the **Axiom of Regularity**.

We caution that this is a rather weak theory; it contains no axioms for generating sets. If M is *any* non-empty transitive set, then

$$(M, \in) \models T_0 .$$

We assign to certain finite sets A a formula, $\psi_A(x_0)$, that completely specifies the set. For example, to the empty set we assign the formula

$$(\forall x_1(\neg(x_1\hat{\in}x_0))) .$$

The sets that we are interested in are finite ordinals and finite sequences of ordinals. We first handle the finite ordinals. We proceed by induction. We have just defined ψ_0. Suppose that k is a finite ordinal and that ψ_i is defined for $i \leq k$. Let $n = 2k + 2$, so that, by induction n is the least natural number such that x_n does not occur in the formula ψ_k. Then ψ_{k+1} is the formula

$$(\forall x_n((x_n\hat{\in}x_0) \leftrightarrow (\psi_k[x_n : x_0] \vee (\exists x_{n+1}(\psi_k[x_{n+1} : x_0] \wedge (x_0\hat{\in}x_{n+1})))))) .$$

Now suppose that A is the ordered pair (i, j), where i and j are finite ordinals. It is routine, though cumbersome, to define $\psi_{(i,j)}$. First we define $\psi_{\{i\}}$ and $\psi_{\{i,j\}}$. Let $n = 2 + 2 \cdot \max\{i, j\}$. As above, n is the least natural number such that x_n does not occur in either ψ_i or ψ_j. Then $\psi_{\{i\}}$ is the formula

$$(\forall x_n((x_n\hat{\in}x_0) \leftrightarrow \psi_i[x_n : x_0])) ,$$

and $\psi_{\{i,j\}}$ is the formula

$$(\forall x_n((x_n\hat{\in}x_0) \leftrightarrow (\psi_i[x_n : x_0] \vee \psi_j[x_n : x_0]))) .$$

Also $\psi_{(i,j)}$ is the formula

$$(\forall x_{n+1}((x_{n+1}\hat{\in}x_0) \leftrightarrow (\psi_{\{i\}} \vee \psi_{\{i,j\}}))) .$$

Finally suppose that A is a function

$$A : i + 1 \mapsto j + 1,$$

where i and j are finite ordinals. Let n be the least natural number such that x_n does not occur in any of the formulas $\psi_{(k,A(k))}$ where $k < i + 1$. For each

$k < i + 1$, let φ_k be the formula

$$\psi_{(k,A(k))}[x_n : x_0] .$$

Then ψ_A is the formula

$$(\forall x_n((x_n \widehat{\in} x_0) \leftrightarrow (\dots((\varphi_0 \vee \varphi_1) \vee \varphi_2) \dots \vee \varphi_i))) .$$

We need to define the *length* $\ell(A)$ of a proof. This we define in the obvious fashion. Suppose that ψ is a formula and that $\psi = \langle m_0, \dots, m_k \rangle$. Then

$$\ell(\psi) = \operatorname{dom} \psi = k + 1 .$$

Suppose that $\langle \psi_0, \dots, \psi_k \rangle$ is a sequence of formulas. Then

$$\ell(\langle \psi_0, \dots, \psi_k \rangle) = \sum_{i=0}^{k} \ell(\psi_i) .$$

We make the following observation. Suppose that A is a proof and $\ell(A) \leq n$. The formulas of A can involve at most n variables and so, by substituting, if necessary, we can suppose that all the variables occurring in the formulas of A are included in the set $\{x_0, \dots, x_n\}$. It follows that A can be coded in a natural fashion by a binary sequence of length at most

$$n[\log_2(n)] ,$$

where $[x]$ is the greatest integer function.

We now define the Gödel sentences from which we shall obtain our formula. This requires the construction of formulas that express many of the concepts we have defined. The explicit construction of these sentences is a tedious affair, which we shall, for the most part, spare the reader. For example we have defined an ordered pair. A formula that expresses this is

$$\begin{aligned}&\exists x_1 \exists x_2 \exists x_3 \exists x_4((x_3 \widehat{\in} x_0) \wedge (x_4 \widehat{\in} x_0) \wedge (x_1 \widehat{\in} x_3) \wedge (x_1 \widehat{\in} x_4) \wedge \\ &(x_2 \widehat{\in} x_4) \wedge (\forall x_5(((x_5 \widehat{\in} x_3) \rightarrow (x_5 \widehat{=} x_1)) \wedge \\ &((x_5 \widehat{\in} x_4) \rightarrow ((x_5 \widehat{=} x_1) \vee (x_5 \widehat{=} x_2)))))) ,\end{aligned}$$

where we have dropped some parentheses and used our abbreviations.

We indicate these formulas as follows. The formula above asserts that 'x_0 is an ordered pair'. We denote this formula by

“$\boxed{x_0}$ is an ordered pair” .

Similarly

“$\boxed{x_3}$ is a function”

indicates the formula we could write down (with painful effort) for the definition of a function. This formula would involve the formula above for specifying ordered pairs. Of course there is no unique way of constructing these formulas, but there are natural formulations based on the definitions we have indicated.

Similarly it is straightforward to find formulas for

$$\text{“}\boxed{x_0}\text{ is a transitive set”},$$

$$\text{“}\boxed{x_0}\text{ is a finite ordinal”},$$

and for

$$\text{“}\boxed{x_0}\text{ is a finite sequence”}.$$

One formula we shall need is a formula which expresses

$$\text{“}\boxed{x_0}\text{ is a finite ordinal, }\boxed{x_1}\text{ is a finite ordinal and }\boxed{x_1} = 2^{\boxed{x_0}}\text{”}.$$

Such a formula may be specified using the previous formulas by formalizing the usual definition; $m = 2^n$ if there exists a sequence s of length $n+1$ such that $s(0) = 0$, such that, for all $k < n$, $s(k+1) = 2 \cdot s(k)$, and such that $m = s(n)$. The axioms of T_0 prove that this uniquely specifies 2^n.

This formula in turn requires a formula expressing

$$\text{“}\boxed{x_0}\text{ is a finite ordinal, }\boxed{x_1}\text{ is a finite ordinal and }\boxed{x_1} = 2 \cdot \boxed{x_0}\text{”},$$

which can be defined in an analogous fashion, noting that one can easily specify a formula expressing

$$\text{“}\boxed{x_0}\text{ is a finite ordinal, }\boxed{x_1}\text{ is a finite ordinal, and }\boxed{x_1} = \boxed{x_0}+2\text{”}.$$

From these formulas we easily obtain a formula which expresses

$$\text{“}\boxed{x_0}\text{ is a finite ordinal, }\boxed{x_1}\text{ is a finite ordinal, and }\boxed{x_1} = \mathrm{Exp}_{2011}(\boxed{x_0})\text{”}.$$

We continue the discussion of the formulas that we shall need.

Perhaps more confusing at first is the formula

$$\text{“}\boxed{x_0}\text{ is a variable”},$$

or the formula,

$$\text{“}\boxed{x_0}\text{ is the variable }x_4\text{”}.$$

The latter is simply the formula

$$\text{“}\boxed{x_0}\text{ is the finite ordinal 11”},$$

which can be taken to be ψ_{11}.

We discuss some formulas that are perhaps a little more difficult to construct. We first consider the formula

$$\text{“}\boxed{x_0}\text{ is a formula”}.$$

For the construction of this formula we define the notion that a set is a formula witness. A set A is a *formula witness* if A is a finite sequence $A =_{\text{.}} \langle a_0, \ldots, a_k \rangle$ such that a_0 is an atomic formula and such that, for each $i + 1 \leq k$, a_{i+1} is an atomic formula or a_{i+1} is obtained from elements of $\{a_0, \ldots, a_i\}$ by a single application of one of the rules of formation. Thus

$$\text{“}\boxed{x_0}\text{ is a formula”}$$

is the formula

$$(\exists x_1(\text{“}\boxed{x_1}\text{ is a formula witness”} \wedge \text{“}\boxed{x_0}\text{ occurs in }\boxed{x_1}\text{”})).$$

From this formula it is routine to build the formulas indicated by

$$\text{“}\boxed{x_0}\text{ is a logical axiom”},$$

indicated by

$$\text{“}\boxed{x_0}\text{ is a sentence in the theory } T_0\text{”},$$

and indicated by

$$\text{“}\boxed{x_0}\text{ is a proof from } T_0\text{”}.$$

The purpose of this formalization of our formalization (and this is the essence of Gödel sentences in general) is that we can build sentences which refer to their own provability.

We first give the standard example of a Gödel sentence modified to our context (in the language $\mathcal{L}(\widehat{=}, \widehat{\in})$ and relative to the theory T_0).

Let $\Phi(x_0)$ be the formula:

$$\text{“}\boxed{x_0}\text{ is a formula with only one free variable, } x_0\text{”}$$

$$\wedge(\exists x_1(\exists x_2(\text{“}\boxed{x_2}\text{ is the sentence } (\neg(\exists x_0(\boxed{x_0} \wedge \psi_A))) \text{ where } A = \boxed{x_0}\text{”}$$

$$\wedge\ \text{“}\boxed{x_2}\text{ is a proof of the sentence }\boxed{x_1}\text{ from the theory } T_0\text{”}))$$

Let Θ be the formula ψ_A, where A is the formula Φ. Let Ω be the sentence

$$(\exists x_0(\Phi \wedge \Theta)).$$

The sentence Ω is a Gödel sentence. In essence Ω asserts that its negation, $(\neg\Omega)$, can be proved from the theory T_0.

By the usual arguments it follows (within our universe of sets) that the theory T_0 does not prove Ω and T_0 does not prove $(\neg\Omega)$; i.e., the sentence Ω is *independent* of the theory T_0. We give the argument.

For trivial reasons, T_0 cannot prove Ω. This is because

$$(\{\emptyset\}, \in) \models T_0$$

and clearly

$$(\{\emptyset\}, \in) \models (\neg\Omega).$$

Therefore we have only to show that $T_0 \nvdash (\neg\Omega)$. Assume toward a contradiction that $T_0 \vdash (\neg\Omega)$. Let M be any transitive set containing such a proof and closed under the necessary witness, so that $(M,\in) \models$ "$T_0 \vdash (\neg\Omega)$". Then $(M,\in) \models \Omega$, but $(M,\in) \models T_0$, and so $(M,\in) \models (\neg\Omega)$, which is a contradiction.

The sentence Ω is too pathological even for our purposes; a proof of $(\neg\Omega)$ cannot belong to a model of T_0 with any *reasonable* extent beyond the length of the proof. The sentences we seek are obtained by two modifications of the formula Φ.

Let $\Phi^*(x_0, x_1)$ be the formula:

$$\text{“}\boxed{x_0}\text{ is a formula with only two free variables, } x_0 \text{ and } x_1\text{”}$$

$$\wedge(\exists x_2(\exists x_3(\text{“}\boxed{x_3}\text{ is the sentence } (\neg(\exists x_1((\exists x_0(\boxed{x_0} \wedge \psi_A))$$

$$\wedge(\exists x_0((x_0 \mathrel{\hat{=}} x_1) \wedge \psi_B))))) \text{ where } A = \boxed{x_0} \text{ and } B = \boxed{x_1}\text{”}$$

$$\wedge\ \text{“}\boxed{x_2}\text{ is a proof of the sentence }\boxed{x_3}\text{ from the theory } T_0\text{”}$$

$$\wedge\text{“}\boxed{x_1}\text{ is a finite ordinal”}$$

$$\wedge(\exists x_4(\text{“}\boxed{x_4} = 10^{24k} \text{ where } k = \boxed{x_1}\text{”}$$

$$\wedge\Psi \wedge \text{ “the length of }\boxed{x_2}\text{ is less than }\boxed{x_4}\text{”})))))\,.$$

The formula $\Psi(x_4)$ shall be defined below. Informally Ψ asserts that there exists a *fragment* of the universe which is reasonable.

Let Θ^* be the formula ψ_A, where A is the formula Φ^*. Let $\Omega(x_1)$ be the formula

$$(\exists x_0(\Phi^* \wedge \Theta^*))$$

and, for each positive integer k, let Ω_k be the sentence,

$$(\exists x_1(\Omega \wedge (\exists x_0((x_0 \mathrel{\hat{=}} x_1) \wedge \psi_k))))\,.$$

Roughly (ignoring the effect of Ψ), the sentence Ω_k asserts that there is a proof from T_0 of $(\neg\Omega_k)$ of length less than n, where $n = 10^{24k}$. The purpose of the function $f(k) = 10^{24k}$ is to ensure that the length of Ω_k is small relative to the number n.

The indicated function, $f(k) = 10^{24k}$, is somewhat arbitrary and can be replaced by any reasonable function of sufficient growth.

We now define the formula $\Psi(x_4)$. Unfortunately this will seem rather technical even when presented in our informal manner for specifying formulas.

We fix some more notation. Suppose that X is a set. Then

$$\mathcal{P}(X) = \{Y \mid Y \subset X\}\,,$$

and thus the formula

$$\text{“}x_1 = \mathcal{P}(x_0)\text{”}$$

asserts that x_1 is the *powerset* of x_0.

Let $\Psi(x_4)$ be the formula:

$$\begin{array}{l}
(\exists x_5\text{“}\,\boxed{x_5}\text{ is a 1001-tuple }\langle A_0,\ldots,A_{1000}\rangle\text{ where}\\
A_0 = \mathrm{Exp}_{1001}(\boxed{x_4});\text{for each }i < 1001, A_{i+1} = \mathcal{P}(A_i);\\
\text{and the standard model corresponding to }A_{1000}\\
\text{satisfies each instance of the axiom of comprehension}\\
\text{which belongs to the set }A_4\text{”})
\end{array}$$

We expand this further:

$$\begin{array}{l}
(\exists x_5(\exists x_6(\exists x_7(\exists x_8\text{“}\,\boxed{x_5}\text{ is a 1001-tuple }\langle A_0,\ldots,A_{1000}\rangle\\
\text{such that }A_0 = \mathrm{Exp}_{1001}(\boxed{x_4});\\
\text{for each }i < 1001, A_{i+1} = \mathcal{P}(A_i); A_4 = \boxed{x_8}\text{ and }A_{1000} = \boxed{x_6}\text{”}\\
\wedge\text{“}\,\boxed{x_7}\text{ is a function }I : S \times M^m \to \{0,1\}\text{ , where }S\\
\text{is the set of formulas which belong to }\boxed{x_8}, m = A_0, M = \boxed{x_6},\\
\text{and }I\text{ is an oracle for }((M,\in),S)\text{ such that }I(\psi,b) = 1\\
\text{for all }b \in M^m\text{ and for all formulas }\psi\text{ such that}\\
\psi\text{ is an instance of the axiom of comprehension.”}))))
\end{array}$$

Thus Ω_k asserts that there is a proof from T_0 of $(\neg\Omega_k)$ of length less than 10^{24k} and further that there exists a *reasonable* transitive set to which this proof belongs. The formula Ψ specifies the exact nature of this transitive set. There is no real reason for our particular choice; one could quite easily modify it, perhaps requiring that the transitive set be larger relative to n and satisfy more sentences.

In our universe of sets $(\neg\Omega_k)$ is true and so there is a proof of $(\neg\Omega_k)$ from T_0. It is not clear just how short such a proof can be. This is a very interesting question. The witness for armageddon is a proof of $(\neg\Omega_k)$ from T_0 of length less than 10^{24k}.

We now make a connection with the title of this chapter. It refers to the game called *The Tower of Hanoi.* One description of this game, redolent of a bygone age, is the following,[1] where the game is called the *Tower of Bramah*:

> In the great temple at Benares, beneath the dome which marks the centre of the world, rests a brass-plate in which are fixed three diamond needles, each a cubit high and as thick as the body of a bee. On one of these needles, at the creation, God placed sixty-four discs of pure gold, the largest disc resting on the brass plate, and the others getting smaller and smaller up to the top one. This is the Tower of Bramah. Day and night unceasingly the priests transfer the discs from one diamond needle to another according to the fixed and immutable laws of Bramah, which require that the priest must not move more than one disc at a time and that he must place this disc on a needle so that there is no smaller disc below it. When the sixty-four discs shall have been thus transferred from the needle on which at the creation God placed them to one of the other needles, tower, temple,

and Brahmins alike will crumble into dust, and with a thunderclap the world will vanish.

The number of separate transfers of single discs which the Brahmins must make to effect the transfer of the tower is $2^{64} - 1$, that is,

$$18\,446\,744\,073\,709\,551\,615\,.$$

There is an analogy between the Tower of Hanoi and our construction.

7 Evidence

We briefly investigate, for various k, the possibilities for evidence that there is no proof of $(\neg\Omega_k)$ from T_0 of length less than 10^{24k}. We are interested in the possibilites for evidence which is not necessarily a proof in the usual sense.

We began with the nonstandard case. Here ZFC denotes the Zermelo–Fraenkel Axioms for set theory together with the *Axiom of Choice.* These axioms are the formal versions of the axioms indicated in §2.

Suppose that $\mathcal{M} = (M, E)$ is a model of ZFC. We suppose that M is a countable set.

Fix $a \in M$ such that

$$\mathcal{M} \models \text{“}a \text{ is a finite ordinal”}$$

and such that a is nonstandard; that is, the set $\{x \in M \mid x\, E\, a\}$ is infinite. Let $b \in M$ be such that

$$\mathcal{M} \models \text{“}b \text{ is the formula } \psi_k, \text{ where } k = a\text{”}\,.$$

Suppose that $c \in M$, that

$$\mathcal{M} \models \text{“}c \text{ is a finite ordinal”}\,,$$

and that $(a, c) \in E$; that is, $\mathcal{M} \models a < c$.

For each $x \in M$ let

$$\tilde{x} = \{y \in M \mid y\, E\, x\}\,.$$

We iterate and define

$$\tilde{\tilde{x}} = \{\tilde{y} \mid y\, E\, x\}\,.$$

Thus, if $x \in M$, $y \in M$, and if $\mathcal{M} \models$ “$x \subset y \times y$” , then $\tilde{\tilde{x}} \subset \tilde{y} \times \tilde{y}$.

Suppose that $A \subset \tilde{c} \times \tilde{c}$ is a set which is *internal* to $\mathcal{M}$; that is, such that $A = \tilde{\tilde{x}}$ for some $x \in M$. We say that A is *consistent* with Ω_a if there exists a model

$$\mathcal{M}^* = (M^*, E^*)$$

such that the following hold:

(1) $\{a, c\} \cup \tilde{a} \cup \tilde{c} \subset M^*$;

(2) $E \cap (\{a, c\} \cup \tilde{a} \cup \tilde{c})^2 = E^* \cap (\{a, c\} \cup \tilde{a} \cup \tilde{c})^2$;

(3) $b \in M^*$, $\tilde{\tilde{b}} \subset M^*$, and $E \cap (\tilde{\tilde{b}})^2 = E^* \cap (\tilde{\tilde{b}})^2$;

(4) $\mathcal{M}^* \models$ "b is the formula ψ_k where $k = a$" ;
(5) there exists $x \in M^*$ such that $\tilde{\tilde{x}} = A$;
(6) $\mathcal{M}^* \models T_0$;
(7) $\mathcal{M}^* \models \Omega[a]$.

Generally one is interested in the case where c is relatively large relative to a (closer to 10^{24a}), in which case (3) and (4) can be achieved by simply modifying the choice of A slightly.

We can now state more precisely our questions. These kinds of questions have been considered in a slightly different context in (Solovay 1989).

1) Suppose that $\mathcal{M}$ and $a \in M$ are as above. Is Ω_a consistent with A for each set $A \subset \tilde{c} \times \tilde{c}$ which is internal to $\mathcal{M}$, where $c = 10^{12a}$?

2) Suppose that $\mathcal{M}$ and $a \in M$ are as above. Is Ω_a consistent with A for each set $A \subset \tilde{c} \times \tilde{c}$ which is internal to $\mathcal{M}$, where $c = 10^{24a}$?

3) Suppose that $\mathcal{M}$ and $a \in M$ are as above. Is Ω_a consistent with A for each set $A \subset \tilde{c} \times \tilde{c}$ which is internal to $\mathcal{M}$, where $c = 10^{24a \cdot 24a}$?

4) Do the answers to (1)–(3) depend on the choice of $\mathcal{M}$ and a?

A positive answer to any of these questions would likely involve new insights to the fundamental questions of computational complexity.

One can also ask whether the answers to these questions are affected by modifying the choice of the critical formula $\Psi(x_4)$ used in the construction of $\Omega(x_1)$. For example, one might consider the extreme case where $\Psi(x_4)$ is trivial (for example the formula $(x_4 \hat{=} x_4)$).

The next theorem answers versions of our questions where c is somewhat small relative to 10^{24a}, for example when

$$c = 10^{[24a/\log a]}.$$

Theorem 7.1 *Suppose that $\mathcal{M} = (M, E)$ is a model of* ZFC, *that $a \in M$, and that*

$$\mathcal{M} \models \text{"}a \text{ is a finite ordinal"}.$$

Suppose that a is nonstandard, and let $c \in M$ be such that, for each k,

$$\mathcal{M} \models \text{"}c^k < 10^{24a}\text{"}.$$

Then A is consistent with Ω_a for each set $A \subset \tilde{c} \times \tilde{c}$ such that A is internal.

It would be interesting to study models $\mathcal{M}$ of T_0 such that

$$\mathcal{M} \models (\exists x_1 \Omega),$$

or more generally to study the mathematics of inconsistency through an analysis of other nonstandard examples.

A striking development of modern set theory is the realization that if, for example, measurable cardinals are consistent, then one has a very rich structure theory for the universe of sets which answers many of the classical questions of the subject. Of course the consequences of the existence of large cardinals leads to a far richer theory.

A typical question involving models of T_0 in which the sentence $(\exists x_1 \Omega)$ holds concerns collateral effects. One instance of this category of questions is the following. Let $\Omega_{\mathrm{ZFC}}(x_0, x_1)$ be a formula which expresses

"$\boxed{x_0}$ is a proof from ZFC of $(\exists x_0(\neg(x_0 \hat{=} x_0)))$ of length less than $\boxed{x_1}$" .

Thus Ω_{ZFC} asserts that the theory ZFC is not consistent and that this is witnessed by a proof with an upper bound on the length.

5) Suppose that $\mathcal{M} = (M, E)$ is a model of T_0 such that $\mathcal{M} \models (\exists x_1 \Omega)$. Let $a \in M$ be least such that $\mathcal{M} \models \Omega[a]$. Must there exist $p \in M$ such that $\mathcal{M} \models \Omega_{\mathrm{ZFC}}[p, b]$ where $b^2 = a$?

8 The standard case, a second sentence

We now consider the standard case. We fix k, and set $n = 10^{24k}$. Transcribed to this setting our first question asks in effect if there is evidence of order $\sqrt{n}$ that Ω_k is false; the second question asks if there is evidence of order n, and, roughly, the third question asks if there is evidence of order $n^{\log n}$.

There is a variety of ways to formulate these questions in the standard case. For example, suppose that $\varphi(x_0)$ is a Σ_0 formula and that the sentence $(\exists x_0 \varphi(x_0))$ is true. Suppose also that there exists a set $A \subset n \times n$ such that $\varphi_0[A]$ holds. Is there a proof of $(\neg\Omega_k)$ from $T_0 \cup \{(\exists x_0 \varphi(x_0))\}$ of length less than n? (of length less than $\sqrt{n}$?).

To avoid some trivialities one needs to assign a weight to the specified sentence $(\exists x_0 \varphi(x_0))$ related to the number of steps required to verify $\varphi[A]$ and ask if there is a proof of $(\neg\Omega_k)$ from $T_0 \cup \{(\exists x_0 \varphi(x_0))\}$ such that

$$\text{"length of proof"} \leq \pi(n, \text{"weight"})$$

for various choices of the function $\pi(x, y)$.

We shall not discuss this further. Instead we shall define and briefly discuss another sentence which will generalize Ω_1. The property we would like Ω_1 to have is that there is no *evidence* of order 10^{24} (or more) that the sequence corresponding to Ω_1 does not exist; the sequence being a proof of $(\neg\Omega_1)$ of length less than 10^{24}. It may be that Ω_1 already has this feature; for the nonstandard versions of Ω_k this is implied by a positive answer to our second or third questions.

We define a sentence $\widetilde{\Omega}_1$. The precise definition requires discussion of machine architecture, and therefore we shall be somewhat vague. Suppose that φ is a sentence of $\mathcal{L}(\hat{=}, \hat{\in})$. The sentence φ is *feasibly true* if φ is *verified* by some machine acting on an input s, where s is a finite binary sequence. One constrains the type of machine allowed and the running time to obtain an arguably realistic notion. Having made these choices one can produce a formula

$$\Psi_{\mathsf{feas}}(x_1) = \text{"}\boxed{x_1}\text{ is a sentence of } \mathcal{L}(\hat{=}, \hat{\in}) \text{ which is feasibly true"} .$$

We avoid the issue of what it means for a machine to verify a sentence by defining a class of *acceptable* machines, and then declaring a sentence to be feasibly true if:

1. The sentence specifies an acceptable machine and asserts that there exists an input for that machine on which the machine halts in a given state after a specified interval of time; suitably transcribed as a statement about sets.

2. The sentence is true.

The choice of $\Psi_{\text{feas}}(x_1)$ then simply depends on definition of an acceptable machine.

For a specific choice of $\Psi_{\text{feas}}(x_1)$ one must argue that one has captured a realistic notion of being 'feasibly true'. It is important to emphasize that there is no requirement that the input for the machine be generated in any feasible manner; the situation is analogous to that for the complexity class NP.

For a given choice of $\Psi_{\text{feas}}(x_1)$ we define $\widetilde{\Omega}_1$.

Let $\Phi^{**}(x_0)$ be the formula:

$$\text{“}\boxed{x_0}\text{ is a formula with only one free variable, } x_0\text{”}$$

$$\wedge(\exists x_1(\exists x_2(\exists x_3(\text{“}\boxed{x_3}\text{ is the sentence } (\neg(\exists x_0(\boxed{x_0}\wedge\psi_A))) \text{ where } A = \boxed{x_0}\text{”}$$

$$\wedge\,\Psi_{\text{feas}}(x_1)$$

$$\wedge\text{“}\boxed{x_2}\text{ is a proof of the sentence }\boxed{x_3}\text{ from the theory } T_0\cup\left\{\boxed{x_1}\right\}\text{”}$$

$$\wedge(\exists x_4(\text{“}\boxed{x_4}=10^{24}\text{”}\wedge\Psi(x_4)\wedge\text{“ the length of }\boxed{x_2}\text{ is less than }\boxed{x_4}\text{”}))))))$$

The formula $\Psi(x_4)$ is the formula indicated in the definition of Φ^{**}. With a reasonable model of the machine, the verification indicated will belong to the set witnessing $\Psi(x_4)$. Otherwise one should modify $\Psi(x_4)$.

Let $\widetilde{\Theta}^*$ be the formula ψ_A, where A is the formula Φ^{**}. Then $\widetilde{\Omega}_1$ is the sentence

$$(\exists x_0(\Phi^{**}\wedge\widetilde{\Theta}^*)).$$

The pathological evidence associated to $\widetilde{\Omega}_1$ has two components. The first is a proof of $(\neg\widetilde{\Omega}_1)$ of length less than 10^{24}. This proof may involve one additional sentence ψ other than the axioms of T_0. The second component is the verification of this sentence. It seems quite likely that the pathology may result from either component separately.

One can easily define a parameter family of these sentences, $\widetilde{\Omega}_k$ for each positive integer k, and then investigate nonstandard solutions. This leads to the analogous versions of our basic questions **1)**–**4)**.

In some sense $\widetilde{\Omega}_1$ is a much better example for supporting our basic thesis. Nevertheless because Ω_1 may in fact have the desired properties, and because Ω_1 is much easier to define, we have chosen to focus on the sentences Ω_k.

9 Final remarks

Is there a sequence of length less than 10^{24} which is a proof of $(\neg\Omega_1)$ from T_0? It is not ridiculous that such a sequence exists, in that we do not have any formal evidence that the sequence does not exist. Much stronger claims can be made (with more general forms of evidence allowed) using $(\neg\tilde{\Omega}_1)$, though in this case the pathological evidence is both a proof of $(\neg\tilde{\Omega}_1)$ and the verification of the additional sentence used.

The existence of such a sequence simply implies that the universe is of *a priori* necessity finite. Our imaginings of large finite sets are a generalization of our experiences which is (in this case) not justified.

The sum total of human experience in mathematics to date; i.e., the number of manuscript pages written to date, is certainly less than 10^{12} pages. With proper inputs and global determination one could verify with current technology that a given sequence is a proof of $(\neg\Omega_1)$ of length less than 10^{24}. The shortest proof from T_0 that no such sequence exists must have length at least 10^{24}. This is arguably beyond the reach of our current experience. The issue of course is the *compression* achieved by the informal style in which mathematical arguments are actually written. One might be able to mitigate this somewhat by using a different formal system in the construction of Ω_1. This again is slightly less of an issue for the case of $(\neg\tilde{\Omega}_1)$, for then one could eliminate evidence derived from formal proofs of length 10^{24} as well as other kinds of evidence of order 10^{24}.

We argue that there are two possibilities.

The first possibility is that there are arbitrarily large finite models of our mathematical experience, where the largeness notion is based on the lengths of proofs or some other reasonable metric. Let $\mathcal{M} = (M, E)$ be a model of ZFC containing a nonstandard finite ordinal. Let $m \in M$ be a finite transitive set in $\mathcal{M}$ such that the submodel $\mathcal{M}_0 = (M_0, E_0)$ is a nonstandard model of our mathematical experience, where

$$M_0 = \{a \in M \mid (a, m) \in E\} = \tilde{m}$$

and

$$E_0 = E \cap (M_0 \times M_0).$$

The relevant parameter (i.e., largeness) of $\mathcal{M}_0$ is some nonstandard finite ordinal, b. Let a be the nonstandard finite ordinal of $\mathcal{M}$ such that

$$10^{24a} \leq b < 10^{24(a+1)},$$

where 10^{24a} and $10^{24(a+1)}$ are calculated in $\mathcal{M}$. Since $a < b$ it follows that $a \in M_0$.

It follows immediately from Theorem 7.1 that there exists a nonstandard model $\mathcal{M}_1 = (M_1, E_1)$ of T_0, such that $\mathcal{M}_1 \models \Omega[a]$ and which contains some portion of the $\mathcal{M}_0$, for example all *subsets* of

$$10^{[24a/\log a]} \times 10^{[24a/\log a]}$$

from that model. Since $\mathcal{M}_1 \models \Omega[a]$, one can define in $\mathcal{M}_1$ (using the witness for $\Psi(x_4)$) a very reasonable model of T_0 containing all subsets of

$$10^{24a} \times 10^{24a}$$

from $\mathcal{M}_1$ (and containing much more). The unfortunate inhabitants of this world have available to them all of the tools which are available to us, and yet have a proof of $(\neg\Omega_a)$ of length less than 10^{24a}. Further they possess all the mathematical evidence of order $10^{[24a/\log a]}$ present in $\mathcal{M}_0$. The force of this argument would be greatly amplified if the answer to either *Question 2* or *Question 3*, at least, is 'yes', for then one could preserve a much larger fragment of $\mathcal{M}_0$. We conjecture that the answer to *Question 3*, at least, is 'yes' and so we conjecture that a much stronger version of this argument is in fact valid.

The second possibility is that there cannot exist arbitrarily large finite models of our mathematical experience. This possibility is in many respects as unfortunate as the possibility that there exists, for some k, a proof of $(\neg\Omega_k)$ of length less than 10^{24k}.

Most mathematicians would argue that as the scale of our mathematical investigations increases so does the depth and beauty of our discoveries, illuminating patterns within patterns ad infinitum; that our collective mathematical vision is not an artifact of the scale of our view but instead it is a glimpse of a world beyond.

10 Appendix

We briefly specify the logical axioms for the language $\mathcal{L}(\hat{=}, \hat{\in})$.

We generalize our conventions and let

$$\varphi[x_i : x_j]$$

be the formula obtained from φ by substituting x_i for each free occurrence of x_j *provided* that each free occurrence of x_j is not within the scope of an occurrence of $\exists x_i$. We no longer require that x_i does not occur in φ. If there is a free occurrence of x_j in φ which is within the scope of an occurrence of $\exists x_i$ then

$$\varphi[x_i : x_j] = \varphi.$$

A formula φ is a *generalization* of ψ if

$$\varphi = (\forall x_{n_1}(\ldots(\forall x_{n_k}\psi)\ldots))$$

for some variables $x_{n_1}, \ldots, x_{n_k}$.

Thus the universal closure of ψ is a generalization of ψ.

Suppose that $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ are formulas and that x_i, x_j are variables. Then all generalizations of the following formulas are logical axioms.

(1) $(\varphi_1 \vee (\neg\varphi_1))$.

(2) $(\varphi_1 \rightarrow (\varphi_1 \vee \varphi_2))$.

(3) $(\varphi_2 \rightarrow (\varphi_1 \vee \varphi_2))$.

(4) $((\varphi_1 \vee \varphi_2) \rightarrow ((\neg\varphi_1) \rightarrow \varphi_2))$.
(5) $((\varphi_1 \rightarrow (\varphi_2 \rightarrow \varphi_3)) \rightarrow ((\varphi_1 \rightarrow \varphi_2) \rightarrow (\varphi_1 \rightarrow \varphi_3)))$.
(6) $(\varphi_1 \rightarrow (\varphi_2 \rightarrow \varphi_1))$.
(7) $((\varphi_1 \rightarrow (\neg\varphi_1)) \rightarrow (\neg\varphi_1))$.
(8) $(\varphi_1 \rightarrow ((\neg\varphi_1) \rightarrow \varphi_2))$.
(9) $((\forall x_j \varphi_1) \rightarrow \varphi_1[x_i : x_j])$.
(10) $((\forall x_i(\varphi_1 \rightarrow \varphi_2)) \rightarrow ((\forall x_i \varphi_1) \rightarrow (\forall x_i \varphi_2)))$.
(11) $(\varphi_1 \rightarrow (\forall x_i \varphi_1))$ if x_i is not a free variable of φ_1.
(12) $(x_i \hat{=} x_i)$
(13) $((x_i \hat{=} x_j) \rightarrow (\varphi_1 \rightarrow \varphi_2))$. if φ_1 is atomic, φ_2 is atomic and if

$$\varphi_1[x_i : x_j] = \varphi_2[x_i : x_j].$$

Notes

1. This description is from de Parville, 1884, as quoted in (Rouse Ball 1905, p. 108).

Bibliography

Rouse Ball, W. W. (1905). *Mathematical recreations and essays.* (4th edn) Macmillan, New York.

Solovay, R. (1989). Injecting inconsistencies into models of PA. *Annals of Pure and Applied Logic*, **44**, 101–132.

Department of Mathematics
University of California
Berkeley, CA 94720-3840
USA
email: woodin@math.berkeley.edu

Complete bibliography

Alexander, J. (1923). A lemma on systems of knotted curves. *Proc. National Academy Sciences*, **9**, 93–5. [Jones]

Andrews, G. (1994). The death of proof? Semi-rigorous mathematics? You've got to be kidding. *The Mathematical Intelligencer*, **16** (4), 16–18. [Effros]

Appel, K. and Haken, W. (1986). The four color proof suffices. *The Mathematical Intelligencer*, **8**, 10–20. [Lolli]

Balaguer, M. (1995). A Platonist epistemology. *Synthèse*, **103**, 303–25. [Maddy, Field]

Baldwin, T. (ed.) (1993). *G. E. Moore: selected writings.* Routledge, London. [Oliveri]

Barwise, J. (1977). An introduction to first-order logic. In *Handbook of mathematical logic* (ed. J. Barwise), pp. 5–46. North-Holland, Amsterdam. [Introduction]

Baumgartner, J. (1973). Ineffability properties of cardinals I. *Colloquia Mathematica Societatus Janós Bolyai*, **10**, 109–30. [Tait]

Benacerraf, P. (1965). What numbers cannot be. *Philosophical Review*, **74**, 47–73. Reprinted in *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 272–94. Cambridge University Press, 1983. [Moschovakis]

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*, **70**, 661–80. Reprinted in *Philosophy of mathematics: selected readings* (2nd edn) (ed. P. Benacerraf and H. Putnam), pp. 272–94. Cambridge University Press, 1983. [Maddy, Martin]

Benacerraf, P. and Putnam, H. (ed.) (1983). *Philosophy of mathematics: selected readings* (2nd edn). Cambridge University Press. [Martin, Introduction]

Bendixson, I. (1883). Quelques theorèmes de la théorie des ensembles de points. *Acta Mathematica*, **2**, 415–29. [Slaman]

Berg, G., Julian, W., Mines, R., and Richman, F. (1975). The constructive Jordan curve theorem. *Rocky Mountain J. Math.*, **5**, 225–36. [Bridges]

Bishop, E. (1967). *Foundations of constructive analysis.* McGraw-Hill, New York. [Bridges]

Bishop, E. (1973). Schizophrenia in contemporary mathematics. *American Math. Soc. Colloquium Lectures.* University of Montana, Missoula. [Bridges]

Bishop, E. and Bridges, D. S. (1985). *Constructive analysis.* Grundlehren der Math. Wissenschaften, Vol. 279. Springer-Verlag, Heidelberg. [Bridges]

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

Boolos, G. and Putnam, H. (1968). Degrees of unsolvability of constructible sets of integers. *Journal of Symbolic Logic*, **33**, 497–513. [Slaman]

Boolos, G. (1984). To be is to be the value of a variable (or some values of some variables). *Journal of Philosophy*, **81**, 430–9. [Field]

Bourbaki, N. (1949). Foundations of mathematics for the working mathematician. *Journal of Symbolic Logic*, **14**, 1–8. [Introduction]

Bridges, D. S. (1981). A constructive proximinality property of finite-dimensional linear subspaces. *Rocky Mountain J. Math.*, **11**, 491–7. [Bridges]

Bridges, D. S. (1994*a*). *Computability: a mathematical sketchbook.* Graduate Texts in Mathematics, Vol. 146. Springer-Verlag, Heidelberg. [Bridges]

Bridges, D. S. (1994*b*). A constructive look at the real number line. In: special issue of *Synthèse* on *Real numbers: generalizations of the reals and theories of continua* (ed. P. Ehrlich), pp. 29–92. [Bridges]

Bridges, D. S. (1998). *Constructive mathematics: a foundation for computable analysis.* To appear in *J. Theoretical Computer Science.* [Bridges]

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

Bridges, D. S., and Demuth, O. (1991). Lebesgue measurability in constructive analysis. *Bull. American Math. Soc.*, **24**, 259–76. [Bridges]

Bridges, D. S. and Ishihara, H. (1990). Linear mappings are fairly well-behaved. *Arch. Math.*, **54**, 558–62. [Bridges]

Bridges, D. S. and Richman, F. (1987). *Varieties of constructive mathematics.* London Math. Soc. Lecture Notes, Vol. 97. Cambridge University Press. [Bridges, Introduction]

Bridges, D. S., Calder, A., Julian, W., Mines, R., and Richman, F. (1982). Picard's Theorem. *Trans. American Math. Soc.*, **269**, 513–20. [Bridges]

Bridges, D. S., Julian, W., and Mines, R. (1989). A constructive treatment of open and unopen mapping theorems. *Zeit. Math. Logik Grundlagen Math.*, **35**, 29–43. [Bridges]

Brouwer, L. E. J. (1907). *Over de grondslagen der wiskunde.* Maas & van Suchtelen, Amsterdam. [Bridges]

Brouwer, L. E. J. (1913). Intuitionism and formalism. *Bull. American Math. Soc.*, **20**, 81–96. Reprinted in *Philosophy of mathematics: selected readings* (2nd edn), (ed. P. Benacerraf and H. Putnam), pp. 77–89. Cambridge University Press. [Introduction]

Budd, M. (1987). Wittgenstein on seeing aspects. *Mind*, **96**, 1–17. [Oliveri]

Buhler, J. (1986). Zero-knowledge proofs. *Focus*, Newsletter of the Mathematical Association of America, **6**, no. 5, October, p. 1. [Lolli]

Burgess, J. (1990). Epistemology and nominalism. In *Physicalism in mathematics* (ed. A. Irvine), pp. 1–150. Kluwer Academic, Dordrecht. [Maddy]

Burton, D. M. (1980). *Elementary number theory.* Allyn and Bacon, Boston, Massachusetts. [Introduction]

Cantor, G. (1883). *Grundlagen einer allgemeinen Mannigfaltigheitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen.* Teubner, Leipzig. English translation: Foundations of the theory of manifolds (trans. U. Parpart), *The Campaigner*, **9** (1976), 60–96. [Tait, Introduction]

Chaitin, G. (1992). *Information, randomness and incompleteness. Papers on algorithmic information theory.* World Scientific, Singapore. [Manin]

Chang, C. C. and Keisler, H. J. (1973). *Model theory.* North-Holland, Amsterdam. [Introduction]

Cheng, H. (1973). A constructive Riemann mapping theorem. *Pacific J. Math.* **44**, 435–54. [Bridges]

Chihara, C. (1973). *Ontology and the vicious circle principle.* Cornell University Press, Ithaca, New York. [Maddy]

Chira, S. (1991). The big test. The week in review. *The New York Times,* 24 March. [Effros]

Church, A. (1936*a*) A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, **1**, 40–1. [Slaman]

Church, A. (1936*b*) An unsolvable problem of elementary number theory. *American Journal of Mathematics*, **58**, 345–63. [Slaman]

Cipra, B. A. (1992). Theoretical computer scientists develop transparent proof techniques. *SIAM News*, **25**, May, p. 1. [Lolli]

Cohen, P. J. (1963–64). The independence of the continuum hypothesis. *Proc. National Academy Sciences*, **50**, 1143–8, and **51**, 105–10. [Introduction]

Cohen, P. J. (1966). *Set theory and the continuum hypothesis.* W. A. Benjamin, New York. [Slaman, Introduction]

Connes, A. (1994). *Noncommutative geometry.* Academic Press, New York. [Effros]

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

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

Dales, H. G. (1998). The mathematician as a formalist. *This volume*, 181–200. [Jones, Prawitz]

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, Introduction]

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. [Dales]

Dauben, J. W. (1979). *Georg Cantor. His mathematics and philosophy of the infinite.* Harvard University Press. Reprinted in paperback by Princeton University Press, 1990. [Introduction]

Davis, M. (1964). Infinite games of perfect information. In *Advances in game theory* (ed. M. Dresher, L. S. Shapley, and A. W. Tucker), pp. 85–101. Annals of Mathematics Studies, Vol. 52. Princeton University Press. [Martin]

Davis, P. and Hersh, M. (1980). *The mathematical experience.* Birkhäuser, Boston. [Dales, Lolli]

Dedekind, R. (1967). Letter to Keferstein. In *From Frege to Gödel: a source book in mathematical logic, 1879–1931* (ed. J. van Heijenoort), pp. 98–103. Harvard University Press, Cambridge, Massachusetts. [George and Velleman]

Detlefsen, M. (1986). *Hilbert's program*, Synthèse Library, Vol. 182. Kluwer Academic, Dordrecht. [Introduction]

Detlefsen, M. (ed.) (1992). *Proof and knowledge in mathematics.* Routledge, London. [Lolli]

Dieudonné, J. (1960). *Foundations of modern analysis.* Academic Press, New York. [Bridges]

Dieudonné, J. (1992). *Mathematics–the music of reason.* Springer-Verlag, Berlin. Translated from *Pour l'honneur de l'esprit humain.* Hachette, Paris, 1987. [Dales]

Dummett, M. A. E. (1959). Truth. *Proceedings of the Aristotelian Society*, new series, **59**, 141–62. Postscript (1974). In *Logic and philosophy for linguists: a book of readings* (ed. J. M. E. Moravcsik), pp. 220–5. Mouton, The Hague. [Martin-Löf]

Dummett, M. A. E. (1973). *Frege: philosophy of language.* Duckworth, London. [Martin-Löf]

Dummett, M. A. E. (1976). What is a theory of meaning? (II). In *Truth and meaning* (ed. G. Evans and J. McDowell), pp. 67–137. Clarendon Press, Oxford. [Martin-Löf]

Dummett, M. A. E. (1977) *Elements of intuitionism.* Clarendon Press, Oxford. [Bridges, Introduction]

Dummett, M. A. E. (1978). The philosophical significance of Gödel's theorem. Reprinted in his *Truth and other enigmas*, pp. 186–201. Harvard University Press, Cambridge, Massachusetts. [George and Velleman, Prawitz]

Dummett, M. A. E.. (1983). The philosophical basis of intuitionistic logic. In *Philosophy of mathematics: selected readings* (2nd edn), (ed. P. Benacerraf and H. Putnam), pp. 97–129. Cambridge University Press. [Introduction]

Dummett, M. A. E. (1987). Reply to Dag Prawitz. In *Michael Dummett: contributions to philosophy* (ed. B. Taylor), pp. 281–6. Martinus Nijhoff, Dordrecht. [Prawitz]

Dummett, M. A. E. (1991). *The logical basis of metaphysics.* Duckworth, London. [Introduction]

Dummett, M. A. E. (1993*a*). Wittgenstein on necessity: some reflections. In *The seas of language*, pp. 446–61. Clarendon Press, Oxford. [Prawitz]

Dummett, M. A. E. (1993*b*). What is mathematics about? Reprinted in his *The seas of language*, pp. 429–45. Clarendon Press, Oxford [George and Velleman]

Dummett, M. A, E. (1994). Reply to Prawitz. In *The philosophy of Michael Dummett* (ed. B. F. McGuinness and G. Oliveri), pp. 292–8. Synthèse Library, Vol. 239. Kluwer Academic, Dordrecht. [Prawitz]

Dummett, M. A. E. (1994). Reply to Oliveri. In *The philosophy of Michael Dummett* (ed. B. F. McGuinness and G. Oliveri), pp. 299–307. Kluwer Academic, Dordrecht. [Dales]

Dummett, M. A. E. (1995). *Frege: philosophy of mathematics*, Duckworth, London. [Introduction]

Ekloff, P. C. (1977). Whitehead's problem is undecidable. *American Math. Monthly*, **83**, 775–88. [Introduction]

Effros, E. (1998). Mathematics as language. *This volume*, 131–45. [Dales]

Enderton, H. (1972). *A mathematical introduction to logic.* Academic Press, New York. [George and Velleman, Introduction]

Esterle, J. R. (1978). Injection de semi-groupes divisibles dans des algèbres de convolution et construction d'homomorphismes discontinus de $C(K)$. *Proc. London Math. Soc.* (3), **36**, 59–85. [Dales]

Feferman, S. (1962). Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, **27**, 259–316. [Field]

Feferman, S. (1988). Hilbert program relativized: Proof theoretical and foundational reduction. *Journal of Symbolic Logic*, **53**, 364–84. [Introduction]

Feferman, S. and Hellman G. (1995). Predicative foundations of arithmetic. *Journal of Philosophical Logic*, **24**, 1–17. [George and Velleman]

Field, H. (1989). *Realism, mathematics and modality.* Blackwells, Oxford. [Field, Maddy]

Field, H. (1991). Metalogic and modality. *Philosophical Studies*, **62**, 1–22. [Field]

Field, H. (1994). Are our logical and mathematical concepts highly indeterminate? In *Midwest studies in philosophy* (ed. P. A. French, T. E. Uehling, Jr., and H. K. Wettstein), Vol. 19, pp. 391–429. [Field]

Field, H. (1998). Which undecidable mathematical sentences have determinate truth values? *This volume*, 291–310. [Dales]

Fowler, P. A. (1988). The Königsberg bridges–250 years later. *The American Mathematical Monthly*, **95**, 42–3. [Lolli]

Frege, G. (1884). *Die Grundlagen der Arithmetik.* W. Koebner, Berlin. Translated as *The foundations of arithmetic* (2nd revised edn) (trans. J. L. Austin). Blackwells, Oxford, 1989. [Introduction, Oliveri]

Frege, G. (1893, 1903). *Grundgesetze der Arithmetic, begriffsschriftlich abgeleitet, I and II.* Hermann Pohle, Jena. Translated as *The basic laws of arithmetic* (trans. M. Furth). University of California Press, Berkeley, 1964. [Introduction]

Frege, G. (1952). On sense and denotation. In *Translations from the philosophical writings of Gottlob Frege* (ed. P. Geach and M. Black). Blackwells, Oxford. [Moschovakis]

Frege, G. (1977). Thoughts. In *Logical investigations*, (ed. P. Geach), pp. 1–30. Blackwells, Oxford. [Introduction]

Friedman, H. M. (1971). Higher set theory and mathematical practice. *Annals of Mathematical Logic*, **2**, 325–57. [Martin]

Friedman, H. (1975). Some systems of second order arithmetic and their use. In *Proceedings of the International Congress of Mathematicians*, Vol. 1, pp. 235–42. Canadian Mathematical Congress. [Slaman]

Friedman, M. (1992). *Kant and the exact sciences.* Harvard University Press, Cambridge, Massachusetts. [Introducton]

Gaifman, H. (1964). Measureable cardinals and constructible sets. *Notices American Math. Soc.*, **11**, 771. [Slaman]

Gale, D. and Stewart, F. M. (1953). Infinite games with perfect information. In *Contributions to the theory of games,* Vol. 2 (ed. H. Kuhn and A. Tucker), pp. 245–66. Annals of Mathematics Studies, Vol. 28. Princeton University Press. [Martin]

Galovich, S. (1989). *Introduction to mathematical structures.* Harcourt Brace Jovanovich, San Diego. [Lolli]

Gelfand, I. M. (1941). Nomierte Ringe. *Rec. Math. N.S.* (*Matem. Sbornik*), **9**, 3–24. [Dales]

Gelfand, I. M. and Naimark, M. A. (1943). On the embedding of normed rings into the ring of operators. *Rec. Math. N. S.* (*Matem. Sbornik*), **12**, 197–213. [Effros]

Gentzen, G. (1934–35). Untersuchungen über das logisches Schliessen. *Mathematische Zeitschrift*, **39**, 176–210, 405–31. [Moschovakis, Prawitz]

Gentzen, G. (1938). Neue Fassung des Wiederspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der Exakten Wissenschaften*, **4**, 19–44. [Prawitz]

Gentzen, G. (1943). Beweisbarkeit und Undbeweisbarkeit von anfangsfällen des transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, **119**, 140-61. [Moschovakis]

George, A. (1987). The imprecision of impredicativity. *Mind*, **96**, 514–18. [George and Velleman]

George, A. (1994). Intuitionism and the poverty of the inference argument. *Topoi*, **13**, 79–82. [George and Velleman]

Gödel, K. (1931).Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatsh. Math. Phys.*, **38**, 173–98. Reprinted as: On formally undecidable propositions of *Principia Mathematica* and related systems I. In *Kurt Gödel: collected works*, Vol. I (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G .H. Moore, R. M. Solovay, and J. van Heijenoort). Oxford University Press, New York, 1986. [Slaman, Introduction]

Gödel, K. (1932). Über Vollständigkeit und Widerspruchsfreiheit. *Ergebnisse eines mathematischen Kolloquium*, **2**, 12–13. Reprinted as: On completeness and consistency. In *Kurt Gödel: collected works*, Vol. 1 (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort), pp. 235–7. Oxford University Press, New York, 1986. [Slaman]

Gödel, K. (1938). The consistency of the axiom of choice and of the generalized continuum hypothesis. *Proc. Nat. Acad. Sci. USA*, **24**, 556–7. [Slaman]

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. [Maddy, Martin, Oliveri, Tait]

Gödel, K. (1965). On undecidable propositions of formal mathematical systems. In *The undecidable. Basic papers on undecidable propositions, unsolvable problems and computable functions* (ed. M. Davis), pp. 39–71. Raven Press, Hewlitt, New York [Slaman]

Gödel, K. (1990). In *Kurt Gödel: collected works*, Vol. 2 (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Soloway, and J. van Heijenoort), p. 403. Oxford University Press, New York. [Oliveri]

Gödel, K. (1995). Ontological proof. In *Kurt Gödel: collected works*, Vol. 3, (ed. S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Soloway, and J. van Heijenoort). Oxford University Press, New York. [Manin]

Goodman, N. D. and Myhill, J. (1978). Choice implies excluded middle. *Zeit. Logik und Grundlagen der Math.* **24**, 461. [Bridges]

Gupta, A. and Belnap, N. (1993). *The revision theory of truth.* MIT Press, Cambridge, Massachusetts. [Field]

Hadamard, J., Baire, R., Borel, E., and Lebesgue, H. (1982). Five letters. Translations in and reprinted in G. Moore. *Zermelo's Axiom of Choice*, pp. 311–320. Springer-Verlag, New York. Originally published in 1905. [Maddy]

Harrington, L. A. (1978). Analytic determinacy and $0^{\#}$. *Journal of Symbolic Logic*, **20**, 685–93. [Slaman]

Hawthorne, J. (1996). Mathematical instrumentalism meets the conjunction objection. *Journal of Philosophical Logic*, **25**, 363-97. [Field]

Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik*, **34**, 879. [Effros]

Heraclitus, fragment 53 DK. In G. Colli, *La sapienza Greca*, Vol. III, p. 34. Eraclito, Milano, 1980. [Introduction]

Hersh, R. (1979). Some proposals for reviving the philosophy of mathematics. *Advances in Mathematics*, **31**, 31–50. [Lolli]

Heyting, A. (1962). After thirty years. In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (ed. E. Nagel, P. Suppes, and A. Tarski), pp. 194–7, Stanford University Press, California. [Bridges, Introduction]

Heyting, A. (1971). *Intuitionism—an introduction* (3rd edn). North-Holland, Amsterdam. [Bridges, Introduction]

Hilbert, D. (1901–1902). Mathematical problems. *Bull. American Math. Soc.* (*N.S.*), **8**, 437–79. [Slaman]

Hilbert, D. (1983). On the infinite. In *Philosophy of mathematics: selected readings* (2nd edn), (ed. P. Benacerraf and H. Putnam), pp. 97–129. Cambridge University Press. [Introduction]

Hildebrandt, S. (1995). *Wahrheit und Wert mathematischer Erkenntnis.* Carl Friedrich von Siemens Stiftung, München. [Manin]

Hodes, H. T. (1980). Jumping through the transfinite: the master code hierarchy of Turing degrees. *Journal of Symbolic Logic*, **45**, 204–20. [Slaman]

Horgan, J. (1993). The death of proof. *Scientific American*, **269**, 74–103. [Effros]

Howard, W. (1980). The formula-as-types notion of construction. In *To H. B. Curry: essays on combinatory logic, lambda calculus and formalism* (ed. J. Seldin and J. Hindley), pp. 479–90. Academic Press, New York. [Tait]

Hurkens, A. J. C., McArthur, M., Moschovakis, Y. N., Moss, L., and Whitney, G. T. (1998). The logic of recursive equations. (To appear.) [Moschovakis]

Isaacson, D. (1987). Arithmetical truth and hidden higher-order concepts. In *Logic Colloquium '85* (ed. Paris Logic Group), pp. 147–69. North-Holland, Amsterdam. [George and Velleman]

Jaffe, A. and Quinn, F. (1993). Theoretical mathematics: toward a cultural synthesis of mathematics and theoretical physics. *Bull. American Math. Soc.*, **29**, 1-13. [Manin]

Jech, T. (1973). *The axiom of choice.* North-Holland, Amsterdam. [Dales, Introduction]

Jech, T. (1978). *Set theory.* Academic Press, New York. [Introduction]

Jensen, R. B. and Solovay, R. M. (1970). Some applications of almost disjoint sets. In *Mathematical logic and foundations of set theory* (ed. Y. Bar-Hillel), pp. 84–104. North-Holland, Amsterdam. [Slaman]

Jockusch, Jr., C. G. (1972). Ramsey's theorem and recursion theory. *Journal of Symbolic Logic*, **37**, 268–80. [Slaman]

Jockusch, Jr., C. G. and Soare, R. I. (1972). Π_1^0 classes and degrees of theories. *Trans. American Math. Soc.*, **173**, 33–56. [Slaman]

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge University Press. [Lolli]

Jones, V. F. R. (1985). A polynomial invariant for knots via von Neumann algebras. *Bull. American Math. Soc.*, **12**, 103–11. [Jones]

Jones, V. F. R. (1990). Knot theory and statistical mechanics. *Scientific American*, **262**, 98–103. [Jones]

Jones, V. F. R. (1998). A credo of sorts. *This volume*, 203–214. [Effros, Prawitz]

Kant, I. (1990). *Critique of pure reason* (trans. N. K. Smith). Macmillan, London. [Oliveri, Introduction]

Kaufman, R. (1984). Fourier transforms and descriptive set theory. *Mathematika*, **31**, 336–9. [Slaman]

Kechris, A. S. (1991). Amenable equivalence relations and the Turing degrees. *Journal of Symbolic Logic*, **56**, 182–94. [Slaman]

Kechris, A. S. (1995). *Classical descriptive set theory.* Springer-Verlag, Heidelberg. [Martin, Slaman]

Kechris, A. S. and Louveau, A. (1987). *Descriptive set theory and the structure of sets of uniqueness.* London Mathematical Society Lecture Note Series, Vol. 128. Cambridge University Press. [Slaman]

Kechris, A. S. and Moschovakis, Y. N. (1978). The Victoria Delfino problems. In *Cabal Seminar 76–77* (ed. A. S. Kechris and Y. N. Moschovakis). Lecture Notes in Mathematics, Vol. 689, pp. 279–82. Springer-Verlag, Heidelberg. [Slaman]

Kechris, A. S. and Woodin, W. H. (1986). Ranks of differentiable functions. *Mathematika*, **33**, 252–78. [Slaman]

Kirkham, R. L. (1995). *Theories of truth.* MIT Press, Cambridge, Massachusetts. [Introduction]

Kleene, S. C. (1936). General recursive functions of natural numbers. *Mathematische Annalen*, **112**, 727–42. [Slaman]

Kleene, S. C. (1952). *An introduction to metamathematics.* North-Holland, Amsterdam and von Nostrand, Princeton. [George and Velleman, Moschovakis, Introduction]

Kleene, S. C. (1955). Hierarchies of number-theoretic predicates. *Bull. American Math. Soc. (N.S.)*, **61**, 193–213. [Slaman]

Kleene, S. C. (1987). Kurt Gödel: 1906–1978. *Biographical memoirs*, **56**, 135–78. [Slaman]

Kline, M. (1972). *Mathematical thought from ancient to modern times.* Oxford University Press, New York. Reprinted in paperback in three volumes, 1990. [Introduction]

Knuth, D. E. (1973). *The art of computer programming. Fundamental algorithms*, Vol. 1, (2nd edn). Addison-Wesley, Reading, Mass. [Moschovakis]

Kreisel, G. (1967). Informal rigour and completeness proofs. In *Problems in the philosophy of mathematics* (ed. I. Lakatos), pp. 138–71. North-Holland, Amsterdam. [Lolli]

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690-716. [Field, Introduction]

Kripke, S. (1980). *Naming and necessity.* Blackwells, Oxford. [Oliveri]

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd edn, enlarged). The University of Chicago Press. [Oliveri]

Kunen, K. (1980). *Set theory, an introduction to independence proofs.* North-Holland, Amsterdam. [Introduction]

Kwiat, P., Weinfurter, H., and Zeilinger, A. (1996). Quantum seeing in the dark. *Scientific American*, **271**, 72–8. [Effros]

Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, **62**, 221–36. [Field]

Levy, A. and Solovay R. M. (1967). Measurable cardinals and the continuum hypothesis. *Israel Journal of Mathematics*, **5**, 234–48. [Maddy]

Linsky, L. (ed). (1952). *Semantics and the philosophy of language*, The University of Illinois Press at Urbana. [Introduction]

Lolli, G. (1987). *La macchina e le dimostrazioni.* Il Mulino, Bologna. [Lolli]

Lolli, G. (1995). *Completeness.* Associazone Italiana Logica e Applicazione, Milan. Preprint. [Lolli]

Louveau, A. and Saint-Raymond, J. (1987). Borel classes and closed games: Wadge-type and Hurewicz-type results. *Trans. American Math. Soc.*, **304**, 431–67. [Martin]

Louveau, A. and Saint-Raymond, J. (1988). The strength of Borel Wadge Determinacy. In *Cabal Seminar 1981–1985* (ed. A. S. Kechris, D. A. Martin, and J. R. Steel), Lecture Notes in Mathematics, Vol. 1333, pp. 1–30. Springer-Verlag, Berlin. [Martin]

Maass, W. (1985). Combinatorial lower bound arguments for deterministic and nondeterministic Turing machines. *Trans. American Math. Soc.*, **292**, 675–93. [Moschovakis]

MacKenzie, D. (1995). The automation of proof: a historical and sociological exploration. *IEEE Annals of the History of Computing*, **17**, 7–29. [Lolli]

Mac Lane, S. (1971). *Categories for the working mathematician.* Springer-Verlag, New York. [Dales]

Maddy, P. (1988*a*). Believing the axioms. I. *Journal of Symbolic Logic*, **53**, 481–511. [Martin]

Maddy, P. (1988*b*). Believing the axioms. II. *Journal of Symbolic Logic*, **53**, 736–64. [Martin]

Maddy, P. (1990). *Realism in mathematics.* Clarendon Press, Oxford. [Dales, Maddy]

Maddy, P. (1993). Does V equal L? *Journal of Symbolic Logic*, **58**, 15–41. [Dales]

Maddy, P. (1995). Naturalism and ontology. *Philosophia Mathematica*, **3**, 248–70. [Maddy]

Maddy, P. (1996). Set-theoretic naturalism. *Journal of Symbolic Logic*, **61**, 490–514. [Maddy]

Maddy, P. (1997). *Naturalism in mathematics.* Clarendon Press, Oxford. [Lolli, Maddy]

Maddy, P. (1998*a*). $V = L$ and Maximize. In *Logic Colloquium '95* (ed. V. Harnik and J. A. Makowsky). (To appear.) [Maddy]

Maddy, P. (1998*b*). Progress in contemporary set theory. In *Mathematical Progress* (ed. H. Breger and E. Grosholz). (To appear.) [Maddy]

Maddy, P. (1998*c*). How to be a naturalist about mathematics. *This volume*, 161–80. [Dales, Field, Lolli]

Manin, Yu (1990). Mathematics as metaphor. In *Proceedings of the International Congress of Mathematicians, Kyoto,* Vol. 2, Mathematical Society of Japan and Springer-Verlag, pp. 1665–71. [Manin]

Manin, Yu. I. (1998). Truth, rigour, and commonsense. *This volume*, 147–59. [Lolli]

Marion, M. (1995). Kronecker's 'Safe haven of real mathematics'. In *Québec studies in the philosophy of science* (ed. M. Marion and A. S. Cohen), pp. 189–215, Kluwer, Dordrecht. [Introduction]

Martin, D. A. (1968). The axiom of determinateness and reduction principles in the analytical hierarchy. *Bull. American Math. Soc.*, **74**, 687–68. [Martin]

Martin, D. A. (1975). Borel determinacy. *Annals of Mathematics*, **102**, 363–71. [Martin]

Martin, D. A. (1998). Mathematical evidence. *This volume*, 215–31. [Dales]

Martin, D. A. (1985). A purely inductive proof of Borel determinacy. In *Proceedings of Symposia in Pure Mathematics*, Vol. 42, pp. 303–8. American Mathematical Society. [Slaman]

Martin, D. A. and Solovay, R. M. (1970). Internal Cohen extensions. *Annals of Mathematical Logic*, **12**, 143–78. [Introduction]

Martin D. A. and Steel, J. R. (1989). A proof of projective determinacy. *Journal American Math. Soc.*, **2**, 71–125. [Martin, Slaman]

Martin-Löf, P. (1971). Hauptzatz for the intuitionistic theory of iterated inductive definitions. In *Proceedings of the 2nd Scandinavian Logic Symposium* (ed. J. E. Fenstad), pp. 179–216. North-Holland, Amsterdam. [Prawitz]

Martin-Löf, P. (1974). *Intuitionistic type theory.* Bibliopolis, Napoli. [Prawitz]

Martin-Löf, P. (1987). Truth of a proposition, evidence of a judgement, validity of a proof. *Synthèse*, **73**, 407–20. [Prawitz]

Martin-Löf, P. (1998). Truth and knowability: on the principles *C* and *K* of Michael Dummett. *This volume,* 105–14. [Introduction]

Matijasevič, Y. (1970). Enumerable sets are diophantine (in Russian). *Doklady Academy Nauk, SSSR*, **191**, 279–82. Translation in *Soviet Math Doklady*, **11** (1970), 354–7. [Slaman]

Mauldin, R. D. (ed.) (1981). *The Scottish book: mathematics from the Scottish café.* Birkhäuser, Boston. [Martin]

McGee, V. (1991). *Truth, vagueness and paradox.* Hackett Publishing Co., Indianapolis. [Field]

Mendelson, E. (1987). *Introduction to mathematical logic.* (3rd edn.) The Wadsworth & Brooks/Cole mathematics series, Belmont, California. [Introduction]

Moore, G. E. (1922). External and internal relations. In *G. E. Moore: selected writings* (ed. T. Baldwin), pp. 79–105. Routledge, London. [Oliveri]

Moore, G. H. (1982). *Zermelo's Axiom of Choice: its origins, development, and influence.* Springer-Verlag, New York. [Dales, Martin, Introduction]

Morris, A. O. (1986). *Linear algebra.* Van Nostrand Reinhold, Wokingham. [Oliveri]

Moschovakis, Y. N. (1980). *Descriptive set theory.* North-Holland, Amsterdam. [Dales, Martin]

Moschovakis, Y. N. (1984). Abstract recursion as a foundation of the theory of algorithms. In *Computation and proof theory* (ed. M. M. Richter *et al.*), Lecture Notes in Mathematics, Vol. 1104, pp. 289–364. Springer-Verlag, Berlin. [Moschovakis]

Moschovakis, Y. N. (1989*a*). The formal language of recursion. *Journal of Symbolic Logic*, **54**, 1216–52. [Moschovakis]

Moschovakis, Y. N. (1989*b*). A mathematical modelling of pure, recursive algorithms. In *Logic at Botik '89*, (ed. A. R. Meyer and M. A. Taitslin). Lecture Notes in Computer Science, Vol. 363, pp. 208–29. Springer-Verlag, Berlin. [Moschovakis]

Moschovakis, Y. N. (1991). A model of concurrency with fair merge and full recursion. *Information and Computation*, **93**, 114–71. [Moschovakis]

Moschovakis, Y. N. (1994*a*). *Notes on set theory.* Undergraduate Texts in Mathematics. Springer-Verlag, New York. [Moschovakis]

Moschovakis, Y. N. (1994*b*). Sense and denotation as algorithm and value. In *Logic Colloquium '90,* Vol. 2, (ed. J. Oikkonen and J. Väänänen). Lecture Notes in Logic, Vol. 2, pp. 210–49. Springer-Verlag, Berlin. [Moschovakis]

Moschovakis, Y. N. (1995). Computable concurrent processes. *Theoretical Computer Science*, **139**, 243–73. [Moschovakis]

Moschovakis, Y. N. (1997). The logic of functional recursion. In *Proceedings of the International Congress for Logic, Methodoloy and the Philosophy of Science*, held in Florence. (To appear.) [Moschovakis]

Moschovakis, Y. N. and Whitney, G. T. (1995). Powerdomains, powerstructures and fairness. In *Computer science logic* (ed. L. Pacholski and J. Tiuryn), Lecture Notes in Computer Science, Vol. 933, pp. 382–96. Springer-Verlag, Berlin. [Moschovakis]

Mumford, D. (1994). Pattern theory: a unifying perspective. *First European Congress of Mathematics* (Paris 1992), Vol. 1, pp. 187–224. Birkhäuser, Basel. [Manin]

Mycielski, J. and Steinhaus, H. (1962). A mathematical axiom contradicting the axiom of choice. *Bulletin de l'Académie Polonaise des Sciences, Série des Sciences Mathématiques, Astronomiques et Physiques*, **10**, 1–3. [Martin]

Nelson, E. (1986). *Predicative arithmetic.* Mathematical Notes, Vol. 32, Princeton. [George and Velleman]

Oliveri, G. (1997*a*). Mathematics. A science of patterns? *Synthèse*, **112**, 379–402 [Oliveri]

Oliveri, G. (1997*b*). Criticism and growth of mathematical knowledge. *Philosophia Mathematica*, (III), **5**, 228–49. [Oliveri, Introduction]

Papineau, D. (1993). *Philosophical naturalism.* Blackwells, Oxford. [Field]

Parsons, C. (1967). Mathematics, foundations of. In *The encyclopedia of philosophy* Vol. 5, (ed. P. Edwards), pp. 188–213. Macmillan, New York. [George and Velleman]

Parsons, C. (1987). Developing arithmetic in set theory without infinity: some historical remarks. *History and Philosophy of Logic*, **8**, 201–13. [George and Velleman]

Parsons, C. (1992). The impredicativity of induction. In *Proof, logic and formalization*, (ed. M. Detlefsen), pp. 139–61. Routledge, London. [George and Velleman]

Pasch, M. (1882). *Vorlesungen über neuere Geometrie.* Teubner, Leipzig. [Lolli]

Penrose, R. (1994). *Shadows of the mind.* Oxford University Press.[Effros, Lolli]

Poincaré, H. (1952). *Science and method* (trans. F. Maitland). Dover, New York. [George and Velleman]

Post, E. L. (1936). Finite combinatory processes. Formulation I. *Journal of Symbolic Logic*, **1**, 103–5. [Slaman]

Prawitz, D. (1965). *Natural deduction: a proof-theoretical study.* Almqvist & Wiksell, Stockholm. [Prawitz]

Prawitz, D. (1971). Ideas and results in general proof theory. In *Proceedings of the 2nd Scandinavian Logic Symposium* (ed. J. E. Fenstad), pp. 235–307. North-Holland, Amsterdam. [Prawitz]

Prawitz, D. (1995). Quine and verificationism. *Inquiry*, **37**, 487–94. [Prawitz]

Putnam, H. (1972). Philosophy of logic. Reprinted in his *Mathematics, matter and method, Philosophical Papers,* Vol. 1 (2nd edn), pp. 323–57. Cambridge University Press, 1979. [Maddy]

Putnam, H. (1980). Models and reality. *Journal of Symbolic Logic*, **45**, 464–82. [Field]

Putnam, H. (1995). Review of Penrose (1994). *Bull. American Math. Soc.* (N.S.), **32**, 370–3. [Lolli]

Quine, W. V. (1961). A basis for number theory in finite classes. *Bull. American Math. Soc.*, **67**, 391–92. [George and Velleman]

Quine, W. V. (1969). *Set theory and its logic.* Harvard University Press, Cambridge, Massachusetts. [George and Velleman]

Quine, W. V. (1970). *Philosophy of logic.* Prentice-Hall, Englewood Cliffs, New Jersey. [Lolli]

Quine, W. V. (1972). Responses. Reprinted in (Quine 1981, pp. 173–86). [Maddy]

Quine, W. V. (1975). Five milestones of empiricism. Reprinted in (Quine 1981, pp. 67–72). [Maddy]

Quine, W. V. (1981). *Theories and things.* Harvard University Press, Cambridge, Massachussetts. [Maddy]

Quine, W. V. (1993). *Pursuit of truth* (revised edn). Harvard University Press, Cambridge, Massachusetts. [Oliveri]

Ramsey, F. P. (1930). On a problem in formal logic. *Proc. London Math. Soc.* (3), **30**, 264–86. [Slaman]

Research Council (1989*a*). *Everybody counts: a report to the nation on the future of mathematics education.* Sciences Education Board, National Research Council, National Academy Press, Washington, D. C. [Effros]

Research Council (1989*b*). *Reshaping school mathematics: a philosophy and framework for curriculum, mathematics.* Sciences Education Board, National Research Council, National Academy Press, Washington, D. C. [Effros]

Responses (1994). Responses to 'Theoretical mathematics etc.', by A. Jaffe and F. Quinn. *Bull. of the American Math. Soc.*, **30**, 161–77. [Manin]

Richards, J. L. (1977). The evolution of empiricism. *British Journal of Philosophy of Science*, **28**, 235–53. [Oliveri]

Richman, F., Bridges, D., Calder, A., Julian, W., and Mines, R. (1982). Compactly generated Banach spaces. *Arch. Math.* **36**, 239–43. [Bridges]

Richman, F. (1983). Church's thesis without tears. *Journal of Symbolic Logic*, **48**, 797-803. [Bridges]

Richman, F. (1996). Interview with a constructive mathematician. *Modern Logic*, **6**, 247–71. [Bridges]

Rips, L. J. (1994). *The psychology of proof.* The MIT Press, Cambridge, Massachusetts. [Lolli]

Rouse Ball, W. W. (1905). *Mathematical recreations and essays.* (4th edn) Macmillan, New York. [Woodin]

Rowbottom, F. (1971). Some strong axioms of infinity incompatible with the axiom of constructibility. *Annals of Mathematical Logic*, **3**, 1–44. [Tait]

Rudin, W. (1970). *Real and complex analysis.* McGraw-Hill, New York. [Bridges]

Rudin, W. (1973). *Functional analysis.* McGraw-Hill, New York. [Bridges]

Russell, B. and Whitehead, A. N. (1910, 1912, 1913). *Principia Mathematica*, Vols. I, II, and III. Cambridge University Press. [Introduction]

Saylor, M. (1997). *Los Angeles Times*, 1 May. [Effros]

Schwichtenberg, H. (1977). Proof theory: some applications of cut-elimination. In *Handbook of mathematical logic* (ed. J. Barwise), pp. 867–95. North-Holland, Amsterdam. [Moschovakis]

Seetapun, D. and Slaman, T. A. (1995). On the strength of Ramsey's theorem. *Notre Dame J. Formal Logic*, **36**, 570–82. [Slaman]

Shapiro, S. (1991). *Foundations without foundationalism: a case for second-order logic.* Clarendon Press, Oxford. [Tait]

Shoenfield, J. R. (1961). The problem of predicativity. In *Essays on the foundations of mathematics* (ed. Y. Bar-Hillel, E. J. J. Poznanski, M. O. Rabin, and A. Robinson), pp. 132–9. Magnes Press, Hebrew University, Jerusalem. [Slaman]

Shoenfield, J. R. (1995). The mathematical work of S. C. Kleene. *Bull. Symbolic Logic*, **1**, 9–43. [Slaman]

Sieg, W. (1988). Hilbert's programme sixty years later. *Journal of Symbolic Logic*, **53**, 338–48. [Introduction]

Simpson, S. G. (1988). Partial realizations of Hilbert's program. *Journal of Symbolic Logic*, **53**, 349–63. [Introduction]

Slaman, T. A. (1998). Mathematical definability. *This volume*, 233–51. [Martin, Tait]

Slaman, T. A. and Steel, J. R. (1988). Definable functions on degrees. In *Cabal Seminar 81–85* (ed. A. S. Kechris, D. A. Martin, and J. R. Steel). Lecture Notes in Mathematics, Vol. 1333, pp. 37–55. Springer-Verlag, Heidelberg. [Slaman]

Smorynski, C. (1977). The incompleteness theorems. In *Handbook of mathematical logic* (ed. J. Barwise), pp. 821–65. North-Holland, Amsterdam. [Introduction]

Solomon, R. (1995). On finite simple groups and their classification. *Notices American Math. Soc.*, **42**, 231–9. [Dales]

Solovay, R. M. (1970). A model of set theory in which every set of reals is Lebesgue measurable. *Annals of Mathematics*, **92**, 1–56. [Dales]

Solovay, R. (1989). Injecting inconsistencies into models of PA. *Annals of Pure and Applied Logic*, **44**, 101–132. [Woodin]

Steel, J. R. (1982). A classification of jump operators. *Journal of Symbolic Logic* **47**, 347–58. [Slaman]

Steen, L. A. (1987). *Calculus today, calculus for a new century.* MAA Notes, Vol. 8, Mathematical Association of America. [Effros]

Stewart, I. (1989). *Galois theory* (2nd edn). Chapman & Hall, London. [Dales]

Swart, E. R. (1980). The philosophical implications of the four-color problem. *American Math. Monthly*, **87**, 697–707. [Lolli]

Tait, W. (1981). Finitism. *Journal of Philosophy*, **78**, 524–46. [Introduction]

Tait, W. (1994). The law of excluded middle and the axiom of choice. In *Mathematics and mind* (ed. A. George), pp. 45–70. Oxford University Press. [Tait]

Tait, W. (1998). Zermelo's conception of set theory and reflection principles. To appear in the proceedings of a conference *Philosophy of Mathematics Today.* [Tait]

Tait, W. (1998). The foundations of set theory. *This volume*, 273–90. [Martin]

Tarski, A. (1952). The semantic conception of truth. In *Semantics and the philosophy of language*, (ed. L. Linsky), pp. 13–47. The University of Illinois Press at Urbana. [Introduction]

Tarski, A. (1956). The concept of truth in formalized languages. In *Logic, semantics, metamathematics*, (ed. A. Tarski). Oxford University Press, pp. 152–278. [Introduction]

Tiuryn, J. (1989). A simplified proof of ddl < dl. *Information and Computation*, **81**, 1–12. [Moschovakis]

Troelstra, A. S., and van Dalen, D. (1988). *Constructivism in mathematics. An introduction*, Vols. I and II. North-Holland, Amsterdam. [Bridges, Introduction]

Turing, A. M. (1936). On computable numbers with an application to the Entscheidungsproblem. *Proc. London Math. Soc.* (3), **42**, 230–65. A correction, **43**, 544–6. [Moschovakis, Slaman]

van Dalen, D. (ed.) (1981). *Brouwer's Cambridge lectures on intuitionism.* Cambridge University Press. [Bridges]

van Stigt, W. P. (1990). *Brouwer's intuitionism.* North-Holland, Amsterdam. [Introduction]

von Neumann, J. (1929). Zur Algebra der Funktionaloperatoren. *Mathematische Annalen*, **102**, 340–427. [Jones]

Wang, H. (1963). Eighty years of foundational studies. Reprinted in his *A survey of mathematical logic*, pp. 34–56. North-Holland, Amsterdam. [George and Velleman]

Weil, A. (1980). *Collected Papers, Vol. 2*, Springer-Verlag, Berlin. [Manin]

Weston, T. (1976). Kreisel, the continuum hypothesis, and second order set theory. *Journal of Philosophical Logic*, **5**, 281–98. [Field]

Wittgenstein, L. (1979*a*). *Notebooks 1914–1916* (2nd edn). Blackwells, Oxford. [Oliveri]

Wittgenstein, L. (1979*b*). *Notes dictated to G. E. Moore in Norway.* In (Wittgenstein 1979*a*), Appendix II, pp. 108–19. [Oliveri]

Wittgenstein, L. (1981). *Tractatus logico-philosophicus.* Routledge, London. [George and Velleman, Oliveri]

Wittgenstein, L. (1983). *Philosophical investigations.* Blackwells, Oxford. [Oliveri]

Woodin, W. H. (1988). Supercompact cardinals, sets of reals and weakly homogeneous trees. *Proceedings of the National Academy of Science USA*, **85**, 6587–91. [Martin]

Woodin, W. H. (1998). The tower of Hanoi. *This volume*, 329–51. [Dales, Jones, Martin]

Wright, C. (1983). *Frege's conception of numbers as objects.* Aberdeen University Press. [Introduction]

Wu, H. (1995). Reviews. *The Mathematical Intelligencer,* **17**, 68–75. [Effros]

Yessenin-Volpin, A. S. (1970). The ultra-intuitionistic criticism and the anti-traditional program for foundations of mathematics. In *Intuitionism and proof theory* (ed. J. Myhill, A. Kino, R. Vesley), pp. 3–46. North-Holland, Amsterdam. [Bridges]

Zeilberger, D. (1993). Theorems for a price: tomorrow's semi-rigorous mathematical culture. *Notices American Math. Soc.*, **40**, 978–81. Reprinted in *The Mathematical Intelligencer*, **17** (4), 11–15. [Effros]

Zermelo, E. (1904). Proof that every set can be well-ordered. In *From Frege to Gödel* (ed. J. van Heijenoort), pp. 139–41. Harvard University Press, Cambridge, Massachusetts, 1967. [Moschovakis]

Zermelo, E. (1908). Untersuchungen über die Grundlagen der Megenlehre I. *Mathematische Annalen* **59**, 261–81. Reprinted as: Investigations in the foundations of set theory I. In *From Frege to Gödel* (ed. J. van Heijenoort), pp. 199–215. Harvard University Press, Cambridge, Massachusetts, 1967. [Moschovakis]

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae*, **16**, 29–47. [Martin, Moschovakis, Tait]

Index