ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS 82

THE FOUNDATIONS OF MATHEMATICS IN THE THEORY OF SETS

J. P. MAYBERRY

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

EDITED BY G.-C. ROTA

Volume 82

The Foundations of Mathematics in The Theory of Sets

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

4 W. Miller, Jr. *Symmetry and separation of variables*
6 H. Minc *Permanents*
11 W. B. Jones and W. J. Thron *Continued fractions*
12 N. F. G. Martin and J. W. England *Mathematical theory of entropy*
18 H. O. Fattorini *The Cauchy problem*
19 G. G. Lorentz, K. Jetter and S. D. Riemenschneider *Birkhoff interpolation*
21 W. T. Tutte *Graph theory*
22 J. R. Bastida *Field extensions and Galois theory*
23 J. R. Cannon *The one-dimensional heat equation*
25 A. Salomaa *Computation and automata*
26 N. White (ed.) *Theory of matroids*
27 N. H. Bingham, C. M. Goldie and J. L. Teugels *Regular variation*
28 P. P. Petrushev and V. A. Popov *Rational approximation of real functions*
29 N. White (ed.) *Combinatorial geometrics*
30 M. Pohst and H. Zassenhaus *Algorithmic algebraic number theory*
31 J. Aczel and J. Dhombres *Functional equations containing several variables*
32 M. Kuczma, B. Chozewski and R. Ger *Iterative functional equations*
33 R. V. Ambartzumian *Factorization calculus and geometric probability*
34 G. Gripenberg, S.-O. Londen and O. Staffans *Volterra integral and functional equations*
35 G. Gasper and M. Rahman *Basic hypergeometric series*
36 E. Torgersen *Comparison of statistical experiments*
37 A. Neumaier *Intervals methods for systems of equations*
38 N. Korneichuk *Exact constants in approximation theory*
39 R. A. Brualdi and H. J. Ryser *Combinatorial matrix theory*
40 N. White (ed.) *Matroid applications*
41 S. Sakai *Operator algebras in dynamical systems*
42 W. Hodges *Model theory*
43 H. Stahl and V. Totik *General orthogonal polynomials*
44 R. Schneider *Convex bodies*
45 G. Da Prato and J. Zabczyk *Stochastic equations in infinite dimensions*
46 A. Bjorner, M. Las Vergnas, B. Sturmfels, N. White and G. Ziegler *Oriented matroids*
47 E. A. Edgar and L. Sucheston *Stopping times and directed processes*
48 C. Sims *Computation with finitely presented groups*
49 T. Palmer *Banach algebras and the general theory of *-algebras*
50 F. Borceux *Handbook of categorical algebra I*
51 F. Borceux *Handbook of categorical algebra II*
52 F. Borceux *Handbook of categorical algebra III*
54 A. Katok and B. Hassleblatt *Introduction to the modern theory of dynamical systems*
55 V. N. Sachkov *Combinatorial methods in discrete mathematics*
56 V. N. Sachkov *Probabilistic methods in discrete mathematics*
57 P. M. Cohn *Skew Fields*
58 Richard J. Gardner *Geometric tomography*
59 George A. Baker, Jr. and Peter Graves-Morris *Padé approximants*
60 Jan Krajicek *Bounded arithmetic, propositional logic, and complex theory*
61 H. Gromer *Geometric applications of Fourier series and spherical harmonics*
62 H. O. Fattorini *Infinite dimensional optimization and control theory*
63 A. C. Thompson *Minkowski geometry*
64 R. B. Bapat and T. E. S. Raghavan *Nonnegative matrices and applications*
65 K. Engel *Sperner theory*
66 D. Cvetkovic, P. Rowlinson and S. Simic *Eigenspaces of graphs*
67 F. Bergeron, G. Labelle and P. Leroux *Combinatorial species and tree-like structures*
68 R. Goodman and N. Wallach *Representations of the classical groups*
69 T. Beth, D. Jungnickel and H. Lenz *Design Theory volume I 2 ed.*
70 A. Pietsch and J. Wenzel *Orthonormal systems and Banach space geometry*
71 George E. Andrews, Richard Askey and Ranjan Roy *Special Functions*
72 R. Ticciati *Quantum field theory for mathematicians*
76 A. A. Ivanov *Geometry of sporadic groups I*
78 T. Beth, D. Jungnickel and H. Lenz *Design Theory volume II 2 ed.*
80 O. Stormark *Lie's Structural Approach to PDE Systems*

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

The Foundations of Mathematics in the Theory of Sets

J. P. MAYBERRY

University of Bristol

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521770347

First published 2000
First paperback edition 2010

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

Library of Congress Cataloguing in Publication Data

Mayberry, John P.
The foundations of mathematics in the theory of sets / J.P. Mayberry.
p. cm (Encyclopedia of mathematics and its applications; v. 82)
Includes bibliographical references and index.
ISBN 0 521 77034 3
1. Set theory. I. Title. II. Series.
QA248.M375 2001
511.3'22-dc21 00-45435 CIP

ISBN 978-0-521-77034-7 Hardback
ISBN 978-0-521-17271-4 Paperback

To Mary Penn Mayberry and Anita Kay Bartlett Mayberry

Contents

Preface

Dancing Master: *All the troubles of mankind, all the miseries which make up history, all the blunders of statesmen, all the failures of great captains – all these come from not knowing how to dance.*
Le Bourgeois Gentilhomme, Act 1, Scene 2

The importance of set-theoretical foundations

The discovery of the so-called "paradoxes" of set theory at the beginning of the twentieth century precipitated a profound crisis in the foundations of mathematics. This crisis was the more serious in that the then new developments in the theory of sets had allowed mathematicians to solve earlier difficulties that had arisen in the logical foundations of geometry and analysis. More than that, the new, set-theoretical approach to analysis had completely transformed that subject, allowing mathematicians to make rapid progress in areas previously inaccessible (in the theory of measure and integration, for example).

All of these advances seemed to be placed in jeopardy by the discovery of the paradoxes. Indeed, it seemed that mathematics itself was under threat. Clearly a retreat to the *status quo ante* was not an option, for serious difficulties once seen cannot just be ignored. But without secure foundations – clear concepts that can be employed without prior definition and true principles that can be asserted without prior justification – the very notion of *proof* is undermined. And, of course, it is the demand for rigorous proof that, since the time of the Greeks, has distinguished mathematics from all of the other sciences.

This crisis profoundly affected some mathematicians' attitudes to their subject. Von Neumann, for example, confessed in a brief autobiographical essay that the existence of the paradoxes of set theory cast a blight on his entire career, and that whenever he encountered technical difficulties in his research he could not suppress the discouraging thought that the problems in the foundations of mathematics doomed the whole mathematical enterprise to failure, in any case.

Mathematics, however, has passed through this crisis, and it is unlikely that a contemporary mathematician would suffer the doubts that von Neumann suffered. Indeed, mathematicians, in general, do not *worry* about foundational questions now, and many, perhaps most, of them are not even interested in such matters. It is surely natural to ask what is the cause of this complacency and whether it is justified.

Of course, every mathematician must master some of the facts about the foundations of his subject, if only to acquire the basic tools and techniques of his trade. But these facts, which are, essentially, just the elements of set theory, can be, and usually are, presented in a form which leaves the impression that they are just definitions or even mere notational conventions, so that their existential content is overlooked. What is more, the exposition of such foundational matters typically begins *in medias res*, so to speak, with the natural numbers and real numbers simply regarded as *given*, so that the beginner is not even aware that these things require proper mathematical definitions, and that those definitions must be shown to be both logically consistent and adequate to characterise the concepts being defined.

These fundamental number systems are nowadays defined using the axiomatic method. But there is a surprisingly widespread misunderstanding among mathematicians concerning the underlying logic of the axiomatic method. The result is that many of them regard the foundations of mathematics as just a branch of mathematical logic, and this encourages them to believe that the foundations of their subject can be safely left in the hands of expert colleagues. But formal mathematical logic itself rests on the same assumptions as do the other branches of mathematics: it, too, stands in need of foundations. Indeed, mathematical logicians are as prone to confusion over the foundations of the axiomatic method as their colleagues.

But this complacency about foundations does have a certain practical justification: modern mathematics does, indeed, rest on a solid and safe foundation, more solid and more safe than most mathematicians realise. Moreover, since mathematics is largely a technical, as opposed to philosophical, discipline, it is not unreasonable that mathematicians should, in the main, get on with the business of pursuing their technical specialities without worrying unduly about foundational questions. But that does not give them licence to pronounce upon matters on which they have not seriously reflected and are ignorant, or to assume that expertise in some special branch of their subject gives them special insight into its foundations.

However, even though it is not, strictly speaking, always necessary for mathematicians to acquire more than a basic knowledge of the foundations of their subject, surely it is desirable that they should do so. Surely the practitioners of a subject the very essence of which is proof and definition ought to be curious about the concepts and principles on which those activities rest.

Philosophers too have an important stake in these questions. Indeed, it is the fundamental role accorded to questions in the philosophy of mathematics that is the characterising feature of western philosophy, the feature that sharply distinguishes it from the other great philosophical traditions.

Problems relating to mathematics and its foundations are to be found everywhere in the writings of Plato and Aristotle, and every major modern philosopher has felt compelled to address them[1]. The subjects that traditionally constitute the central technical disciplines of philosophy – logic, epistemology, and metaphysics – cannot be studied in any depth without encountering problems in the foundations of mathematics. Indeed, the deepest and most difficult problems in those subjects often find their most perspicuous formulations when they are specialised to mathematics and its foundations. Even theology must look to the foundations of mathematics for the clearest and most profound study yet made of the nature of the infinite.

Unfortunately, the complacency, already alluded to, among mathematicians concerning the foundations of their subject has had a deleterious effect on philosophy. Deferring to their mathematical colleagues' technical competence, philosophers are sometimes not sufficiently critical of received opinions even when those opinions are patently absurd.

The mathematician who holds foolish philosophical opinions – about the nature of truth or of proof, for example – is protected from the consequences of his folly if he is prepared to conform to the customs and

[1] This is notoriously the case with Descartes, Leibniz, Kant, and, of course, Frege, who is the founder of the modern analytic school of philosophy; but it is no less true of Berkeley, Hume, and Schopenhauer. Among twentieth century philosophers, Husserl, Russell, and Wittgenstein come to mind.

mores of his professional tribe. But the philosopher who follows him in adopting those opinions does not have that advantage.

In any case, it is one thing to flirt with anarchist views if one lives in a settled, just, and well-policed society, but quite another if one is living in a society in which the institutions of law and justice threaten to collapse. Twice in the last two hundred years mathematicians have been threatened with anarchy – during the early nineteenth century crisis in the foundations of analysis and the early twentieth century crisis in the foundations of set theory – and in both of these crises some of the best mathematicians of the day turned their attention to re-establishing order.

The essential elements of the set-theoretical approach to mathematics were already in place by the early 1920s, and by the middle of the century the central branches of the subject – arithmetic, algebra, geometry, analysis, and logic – had all been recast in the new set-theoretical style. The result is that set theory and its methods now permeate the whole of mathematics, and the idea that the foundations of all of mathematics, including mathematical logic and the axiomatic method, now lie in the theory of sets is not so much a theory as it is a straightforward observation.

Of course that, on its own, doesn't mean that set theory is a *suitable* foundation, or that it doesn't require justification. But it does mean that any would-be reformer had better have something more substantial than a handful of new formalised axioms emblazoned on his banner. And he had better take it into account that even mathematical logic rests on set-theoretical foundations, and so is not available to him unless he is prepared to reform *its* foundations.

The point of view embodied in this book

My approach to set theory rests on one central idea, namely, that the modern notion of *set* is a refined and generalised version of the classical Greek notion of *number* (*arithmos*), the notion of number found in Aristotle and expounded in Book VII of Euclid's *Elements*. I arrived at this view of set theory more than twenty years ago when I first read *Greek Mathematical Thought and the Origin of Algebra* by the distinguished philosopher and scholar Jacob Klein.

Klein's aim was to explain the rise of modern algebra in the sixteenth and seventeenth centuries, and the profound change in the traditional concept of number that accompanied it. But it struck me then with the force of revelation that the later, nineteenth century revolution in the foundations of mathematics, rooted, as it was, in Cantor's new theory of transfinite numbers, was essentially a return to Greek arithmetic as Klein had described it, but in a new, non-Euclidean form.

As Klein points out, in Greek mathematics a number was defined to be a finite plurality composed of units, so what the Greeks called a number (*arithmos*) is not at all like what *we* call a number but more like what we call a set. It is having a finite size (cardinality) which makes a plurality a "number" in this ancient sense. But what is it for a plurality to have a finite size? *That* is the crucial question.

The Greeks had a clear answer: for them a definite quantity, whether continuous like a line segment, or discrete – a "number" in their sense – must satisfy the axiom that *the whole is greater than the part*[2]. We obtain the modern, Cantorian notion of set from the ancient notion of number by abandoning this axiom and acknowledging as finite, in the root and original sense of "finite"– "limited", "bounded", "determinate", "definite" – certain pluralities (most notably, the plurality composed of all natural numbers, suitably defined) which on the traditional view would have been deemed infinite.

By abandoning the Euclidean axiom that the whole is greater than the part, Cantor arrived at a new, *non-Euclidean arithmetic*, just as Gauss, Lobachevski, and Bolyai arrived at *non-Euclidean geometry* by abandoning Euclid's Axiom of Parallels. Cantor's innovation can thus be seen as part of a wider nineteenth century program of correcting and generalising Euclid.

Cantor's non-Euclideanism is much more important even than that of the geometers, for his new version of classical arithmetic that we call *set theory* serves as the foundation for the whole of modern mathematics, including geometry itself. The set-theoretical approach to mathematics is now taken by the overwhelming majority of mathematicians: it is embodied in the mathematical curricula of all the major universities and is reflected in the standards of exposition demanded by all the major professional journals.

Since the whole of mathematics rests upon the notion of set, this view of set theory entails that the whole of mathematics is contained in *arithmetic*, provided that we understand "arithmetic" in its original and historic sense, and adopt the Cantorian version of finiteness. In set theory, and the mathematics which it supports and sustains, we have

[2] This is Common Notion 5 in Book I of the *Elements*.

made real the seventeenth century dream of a *mathesis universalis*, in which it is possible to express the exact part of our thought[3].

But what are the practical consequences of this way of looking at set theory for mathematics and its foundations? They are, I am convinced, profound and far-reaching, both for orthodox set-theoretical foundations, and for the several dissenting and heterodox schools that go under various names – "constructivism", "intuitionism", "finitism", "ultra-intuitionism", etc. – but whose common theme is the rejection of the great revolution in mathematical practice that was effected by Cantor and his followers.

For orthodox foundations the principal benefit of looking at things in this way is that it enables us to see that the central principles – axioms – of set theory are really *finiteness principles* which, in effect, assert that certain multitudes (pluralities, classes, species) are finite in extent and *for that reason* form sets.

Taking finitude (in Cantor's new sense) to be the defining characteristic of sets, as the Greeks took it (in their sense) to be the defining characteristic of numbers (*arithmoi*), allows us to see why the conventionally accepted axioms for set theory – the Zermelo–Fraenkel axioms – are both natural and obvious, and why the unrestricted comprehension principle, which is often *claimed* as natural and obvious (though, unfortunately, self-contradictory), is neither.

This is a matter of considerable significance, for there is a widespread view that all existing axiomatisations of set theory are more or less *ad hoc* attempts to salvage as much of the "natural" unrestricted comprehension principle – the principle that the extension of any well-defined property is a set – as is consistent with avoiding outright self-contradiction[4]. On this view set theory is an unhappy compromise, a botched job at best.

Hence the widespread idea that set theory must be presented as an axiomatic theory, indeed, as an axiomatic theory *formalised* in first order mathematical logic. It is felt that the very formalisation itself somehow confers mathematical respectability on the theory formalised. But this is a serious confusion, based on a profound misunderstanding of the logical and, indeed, *ontological* presuppositions that underlie the axiomatic method, formal or informal.

The mathematician's "set" is the mathematical logician's "domain of discourse", so conventional ("classical") mathematical logic is, like every

---

[3] Perhaps we might more appropriately describe the theory as an *arithmetica universalis*, a *universal arithmetic* which encompasses the whole of mathematics.

[4] See Quine's *Set Theory and its Logic*, for example.

other branch of mathematics, based on set theory[5]. This means, among other things, that we cannot use the standard axiomatic method to establish the theory of sets, on pain of a circularity in our reasoning.

Moreover, on the arithmetical conception of set the totality of all sets, since it is easily seen not to be a set, is not a conventional domain of discourse either. Hence quantification over that non-conventional domain (which is *absolutely infinite* in Cantor's terminology) cannot simply be *assumed* to conform to the conventional, "classical" laws.

As Brouwer repeatedly emphasised, since classical logic is the logic of the finite, the logic of infinite domains must employ different laws. And, of course, in the present context "finite domain" simply means "set". The consequences of this view for the *global* logic of set theory are discussed at length in Section 3.5 and Section 7.2.

But what are the consequences of this *arithmetical* conception of set for those who reject Cantor's innovations – the intuitionists, finitists, constructivists, etc., of the various schools?

Klein's profound scholarship is very much to the point here. For the one thing on which all these schools agree is the central importance of the system of natural numbers as the basic *datum* of mathematics. But Klein shows us that, on the contrary, the natural numbers are a recent invention: the oldest mathematical concept we have is that of *finite plurality* – the Greek notion of *arithmos*. This is so important a matter that I have devoted an entire chapter (Chapter 2) to its dicussion.

When the natural number system is taken as a primary *datum*, something simply "given", it is natural to see the principles of proof by mathematical induction and definition by recursion along that system as "given" as well. We gain our knowledge of these numbers when we learn to count them out and to calculate with them, so we are led to see these *processes* of counting out and calculating as *constitutive* of the very notion of natural number. The natural numbers are thus seen as what we arrive at in the *process* of counting out: $0, 1, 2, \ldots$, where the dots of ellipsis, "...", are seen as somehow self-explanatory – after all, we all know how to continue the count no matter how we have taken it. But those dots of ellipsis contain the whole mystery of the notion of natural number!

If, however, we see the notion of natural number as a secondary

[5] Thus set theory stands the "logicist" view of Frege and Russell on its head: arithmetic isn't a branch of logic, logic is a branch of arithmetic, the non-Euclidean arithmetic of Cantor that we call set theory.

growth on the more fundamental notion of *arithmos* – finite plurality, in the original Greek sense of "finite" – then the principles of proof by induction and definition by recursion are no longer just "given" as part of the raw data, so to speak, but must be established from more fundamental, set-theoretical principles.

Nor are the *operations* of counting out or calculating to be taken as primary data: they too must be analysed in terms of more fundamental notions. We are thus led to reject the *operationalism* that all the anti-Cantorian schools share.

For us moderns numbers take their being from what we can *do* with them, namely count and calculate; but Greek "numbers" (*arithmoi*) were objects in their own right with simple, intelligible natures. Our natural numbers are things that we can (in principle) *construct* (by counting out to them); Greek numbers were simply "there", so to speak, and it would not have occurred to them that their numbers had to be "constructed" one unit at a time[6].

I am convinced that this operationalist conception of natural number is the central fallacy that underlies *all* our thinking about the foundations of mathematics. It is not confined to heretics, but is shared by the orthodox Cantorian majority. This *operationalist fallacy* consists in the assumption that the mere *description* of the natural number system as "what we obtain from zero by successive additions of one" suffices *on its own* to define the natural number system as a unique mathematical structure – the assumption that the operationalist description of the natural numbers is itself what provides us with a *guarantee* that the system of natural numbers has a unique, fixed structure.

Let me not be mistaken here: the existence of a unique (up to isomorphism) natural number system is a *theorem* of orthodox, Cantorian mathematics. The fallacy referred to thus does *not* consist in supposing that *there is* a unique system of natural numbers, but rather in supposing that the existence of this system, and its uniqueness, are immediately given and do not need to be *proved*. And if we abandon Cantorian orthodoxy we thereby abandon the means with which to prove these things.

[6] Oswald Spengler, who thought that the mathematics of a civilization held a clue to its innermost nature, contrasted the *Apollonian* culture of classical Greece, which was static and contemplative, with the *Faustian* culture of modern Europe, which is dynamic and active. Whatever the virtues of his general thesis, he seems to have got it right about the mathematics. The "operationalism" to which I refer here seems to be quintessentially Faustian in his sense, which perhaps explains its grip on our imaginations.

But if we acknowledge that the natural numbers are not *given* to us, the alternative, if we decide to reject Cantor's radical new version of finitude, is to return to arithmetic as practiced by the mathematicians of classical Greece, but equipped now with the more powerful and more subtle techniques of modern set theory. If we should decide to do this we should be going back to the very roots of our mathematical culture, back before Euclid and Eudoxus to its earliest Pythagorean origins. We should have to rethink our approach to geometry and the Calculus. It is a daunting prospect, though an exciting one.

The resulting theory, which I call *Euclidean set theory* by way of contrast with *Cantorian set theory*, the modern orthodoxy, is very like its Cantorian counterpart, except that Cantor's assumption that the species of natural numbers forms a set is replaced by the traditional Euclidean assumption that every set is strictly larger than any of its proper subsets.

This theory, not surprisingly, constitutes a radical departure from Cantorian orthodoxy. But it stands in even sharper contrast to the various operationalist theories which have been put forward as alternatives to that orthodoxy. So far from taking the natural numbers as given, Euclidean set theory forces us to take seriously the possibility that there is no unique natural number system, and that the various ways of attempting to form such a system lead to "natural number systems" of differing lengths.

But *should* we abandon Cantorian orthodoxy? There is obviously a *prima facie* case against the Cantorian account of finiteness, and, indeed, that case was made by some of his contemporaries. But against that there is the experience of more than one hundred years during which Cantor's ideas have been the engine driving a quite astonishing increase in the subtlety, power, and scope of mathematics.

Perhaps I should come clean with the reader and admit that I am attracted to the anti-Cantorian position. I put it no stronger than that because the issue is by no means clear-cut, and we do not yet know enough to be sure that the Cantorian conception of finiteness should be rejected.

Indeed, it seems to me that the common failing of all the advocates of the various alternatives to Cantorian orthodoxy is that they fail to appreciate how simple, coherent, and plausible are the foundational ideas that underlie it. These enthusiasts rush forward with their proposed cures without having first carried out a proper diagnosis to determine the nature of the disease, or even whether there *is* a disease that requires their ministrations.

Accordingly, I shall devote much of my attention to a careful, sympathetic, and detailed treatment of the Cantorian version of the theory. This is of interest in its own right, for this is the theory on which all of current mathematics rests. But it is also essential for those who are dissatisfied (or who fancy themselves dissatisfied) with the current orthodoxy, to discover what principles that orthodoxy really rests on, and to determine exactly where its strengths and weaknesses lie.

I have divided my exposition into four parts. *Part One* deals with the criteria which any attempt to provide foundations for mathematics must meet, and with the significance of the Greek approach to arithmetic for modern foundations.

Part Two is an exposition of the elements of set theory: the basic concepts of set theory, which neither require, nor admit of, definition, but in terms of which all other mathematical concepts are defined; and the basic truths of set theory, which neither require, nor admit of, proof, but which serve as the ultimate assumptions on which all mathematical proofs ultimately rest. The theory presented in *Part Two* is common to both the Cantorian and the Euclidean versions of set theory.

Part Three is an exposition of the Cantorian version of the theory and *Part Four* of the Euclidean. I have also included an appendix which deals with logical technicalities.

This, then, is the point of view embodied in this book: all of mathematics is rooted in *arithmetic*, for the central concept in mathematics is the concept of a plurality limited, or bounded, or determinate, or definite – in short, *finite* – in size, the ancient concept of *number* (*arithmos*).

From this it follows that there are really only two central tasks for the foundations of mathematics:

1. To determine what it is to be *finite*, that is to say, to discover what basic principles apply to finite pluralities *by virtue of their being finite.*
2. To determine what logical principles should govern our reasoning about *infinite* and *indefinite* pluralities, pluralities that are *not* finite in size.

On this analysis, all disputes about the proper foundations for mathematics arise out of differing solutions to these two central problems.

Such a way of looking at things is not easily to assimilate to any of the well-known "isms" that have served to describe the various approaches to the study of mathematical foundations in the twentieth century. But to my mind it has a certain attractive simplicity. Moreover, it is rooted in the history of mathematics and, indeed, takes as its starting point the oldest mathematical concept that we possess.

Acknowledgements

I am especially grateful to Moshé Machover, T.E. Conlon, and Dickon Lush, all of whom read drafts of my manuscript and made many useful suggestions on matters of both style and substance. H.E. Rose and J.C. Shepherdson gave me helpful advice on the material in Chapters 8 and 9.

Alberto Peruzzi invited me to spend a study leave at the University of Florence, during which I started the book, and my many stimulating conversations with him, then and since, have proved invaluable. I shall also always be grateful to him and his family for their wonderful hospitality which made my stay there so pleasant and productive.

My ideas on the concept of *arithmos* were developed in discussions over a number of years with C.J. Rowe and Paul Pritchard, both whom gave me valuable advice in drafting Chapter 2.

I must also acknowledge my debt to my former students Maria del Camino Cañon Loyes, Vincent Homolka, Kevin Lano, Peter Fletcher, Martyn Prigmore, and Nicholas Bamber for many stimulating and fruitful conversations on both technical and philosophical matters.

I am especially indebted to my former student S. J. Popham. The key results in Section 11.2 are due to him, as is the concept of a cumulation model (Definition 11.2.2) upon which they depend. These results are of fundamental importance for the whole approach to foundations taken in Part Four, for they provide crucial information on what *cannot* be proved from the most obvious and natural axioms.

Moreover, it was he who called my attention to the importance of the Ackermann simply infinite system treated in Section 10.6, and who, with the help of a suggestion by Robin Gandy, laid down a definition of that system in the special case of pure sets. Without his work this would have been a very different book.

Finally, I want to thank my wife, Anita, for her help and encouragement. She has given me invaluable advice on the prose composition, and her gentle but persistent criticism has greatly improved the exposition.

Part One

Preliminaries

It is the mark of an educated man to look for precision in each kind of enquiry just to the extent that the nature of the subject allows.

Aristotle

1
The Idea of Foundations for Mathematics

1.1 Why mathematics needs foundations

Mathematics differs from all the other sciences in requiring that its propositions be proved. Certainly no one will deny that proof is the goal of mathematics, even though there may be disagreement over whether, or to what extent, that goal is achieved. But you cannot prove a proposition unless the concepts employed in formulating it are clear and unambiguous, and this means that the concepts used in a proof either must be basic concepts that can be grasped directly and can be seen immediately to be clear and unambiguous, or must be rigorously *defined* in terms of such basic concepts. Mathematics, therefore, since it is about proof is also about definition.

Now definition and proof are both species of the genus *explanation*: to define something is to explain *what* it is; to prove something is to explain *why* it is true. All scholars and scientists, of course, deal in explanation. But mathematicians are unique in that they intend their explanations to be complete and final: that must be their aim and ideal, even if they fail to realise it in full measure. From these simple observations many consequences flow.

Perhaps the most important of them concerns the mathematician's claims to truth. Because he deals in proof, those claims must be absolute and unqualified. Whether they are justified, either in general, or in particular cases, is, of course, quite another matter: but that they are, in fact, made cannot be denied without stripping the word "proof" of all meaning. To claim to have proved something is to claim, among other things, that it is true, that its truth is an objective fact, and that its being so is independent of all authority and of our wishes, customs, habits, and interests. Where there are no truth and falsehood, objectively determined, there can be no proof; and where there is no proof there can be no mathematics.

No doubt all of this is at odds with the *Zeitgeist*[1]: it would seem that we must come to terms with the fact that when there is disagreement about a genuine mathematical proposition, someone must be right and someone must be wrong. But the requirement that we must lay unqualified claims to truth in mathematics is quite compatible with our maintaining a prudent and healthy scepticism about such claims: what it rules out is dogmatic or theoretical scepticism.

You may, as a mathematician, reasonably doubt that such and such a theorem is true, or that such and such a proof is valid: indeed, there are many occasions on which it is your professional duty to do this, even to the point of struggling to maintain doubt that is crumbling under the pressure of argument: for it is precisely when you begin to settle into a conviction that you are most liable to be taken in by a specious but plausible line of reasoning. When your business is judging proofs you must become a kind of professional sceptic. But scepticism, properly understood, is an attitude of mind, not a theory, and you cannot systematically maintain that there is no such thing as a true proposition or a valid argument and remain a mathematician.

A proof, to be genuine, must still all reasonable doubts as to the truth of the proposition proved. But the doubts to be stilled are those that pertain to that proposition: a proof need not, indeed cannot, address general sceptical doubts. Anyone who proposes to pass judgement on the validity of an intended proof must address his attention to the propositions and inferences contained in the argument actually presented. It won't do to object to a particular argument on the ground that all argument is suspect. The fact, for example, that people often make mistakes in calculating sums does not provide grounds for concluding that any particular calculation is incorrect, or even uncertain: each must be judged separately, on its own merits.

In the final analysis, there are only two grounds upon which you may reasonably call the efficacy of a purported proof into question: you may dispute the presuppositions upon which the argument rests, or you may dispute the validity of one or more of the inferences by means of which the argument advances to its conclusion. If, after careful, and perhaps prolonged, reflection, you cannot raise an objection to an argument on

[1] Cantor complained of the "Pyrrhonic and Academic scepticism" that prevailed in his day. *Plus ça change …*

either of these two grounds, then you should accept it as valid and its conclusion as true.

Here we must include among the presuppositions of a proof not only the truth of the propositions that are taken as unproved starting points of the argument, but also the clarity, unambiguity, and unequivocality of the concepts in which the propositions employed in the argument are couched.

Of course in practice, actual proofs start from previously established theorems and employ previously defined concepts. But if we persist in our analysis of a proof, always insisting that, where possible, assertions should be justified and concepts defined, we shall eventually reach the ultimate presuppositions of the proof: the propositions that must be accepted as true without further argument and the concepts that must be understood without further definition. Of course when I say that these things *must* be accepted without proof or understood without definition I mean that they must be so accepted and so understood if the given proof is to be judged valid and its conclusion true.

If we were to carry out such a complete analysis on all mathematical proofs, the totality of ultimate presuppositions we should then arrive at would obviously constitute the foundations upon which mathematics rests. Naturally, I'm not planning to embark on the enterprise of analysing actual proofs to discover those foundations. My point here is rather that solely in virtue of the fact that mathematics is about proof and definition it must of necessity *have* foundations, ultimate presuppositions – unproved assertions and undefined concepts – upon which its proofs and definitions rest.

Of course that observation is compatible with there being a motley of disparate principles and concepts underlying the various branches of the subject, with no overarching ideas that impose unity on the whole. The question thus arises whether it is possible to discover a small number of clear basic concepts and true first principles from which the whole of mathematics can be systematically developed: that is, I suspect, what most mathematicians have in mind when they speak of providing foundations for mathematics.

From the very beginnings of the subject, that is to say, from the time when proof became central in mathematics, mathematicians and philosophers have been aware of the need to provide for foundations in the ideal and general sense just described. But there are particular, and pressing, practical reasons why present day mathematics needs foundations in this sense. Mathematics today is, for mathematicians, radically different from what it was in the relatively recent past, say one hundred and fifty years ago, and, indeed, come to that, from what it is now for professional *users* of mathematics, such as physicists, engineers, and economists. The difference lies in the greatly enhanced role that definition now plays. Present day mathematics deals with rigorously defined mathematical structures: groups, rings, topological spaces, manifolds, categories, etc. Traditional mathematics, on the other hand, was based on geometrical and kinematical intuition. Its objects were idealised shapes and motions. They could be imagined – pictured in the mind's eye – but they could not be rigorously defined.

Now it is precisely in our possession of powerful and general methods of rigorous definition that we are unquestionably superior to our mathematical predecessors. However, this superiority does not consist primarily in our basic definitions being more certain or more secure – although, indeed, they are more certain and secure, as are the proofs that employ them – but rather in the fact that they can be generalised and modified to apply in circumstances widely remote from those in which they were originally conceived.

There is a certain irony here. For although the earliest pioneers of modern rigour – Weierstrass for example – set out in search of safer, more certain methods of definition and argument by cutting mathematics free of its former *logical* dependence on geometrical and kinematical intuitions, they have, paradoxically, enormously enlarged the domain in which those intuitions can be applied.

When we give a rigorous "analytic" (i.e. non-geometrical, non-kinematical) definition of "limit" or "derivative" we do, undoubtedly, attain a greater certainty in our proofs. But, what is just as important, we can *generalise* a rigorous, analytic definition, while a definition based on geometrical or kinematical intuition remains tied to what we can actually visualise. By purging our definitions of their *logical* dependence on geometrical and kinematical intuition, we clear the way for transferring our insights based on that intuition to "spaces", for example, infinite dimensional ones, in which intuition, in the Kantian sense of sensual intuition – images in the mind's eye – is impossible. The mathematicians of the nineteenth century noticed that by a novel use of definition they could convert problems in geometry into problems in algebra and set theory, which are more amenable to rigorous treatment[2]. What they didn't

[2] Descartes saw that problems in geometry could be converted into problems in algebra. But his algebra, the algebra of real numbers, rested logically on geometrical conceptions.

foresee – how could they have foreseen it? – was the enormous increase in the scope of mathematics that these new methods made possible. By banishing "intuitive" ("*anschaulich*") geometry from the logical foundations of mathematics, they inadvertently, and quite unintentionally, gave that geometry a new lease of life.

But it was the technique of axiomatic definition that made the transition from traditional to modern mathematics possible. Naively, an axiomatic definition defines a *kind* or *species* of mathematical structure (e.g. *groups, rings, topological spaces, categories,* etc.) by laying down conditions or *axioms* that a structure must satisfy in order to be of that kind. Axiomatic definition is the principal tool employed in purging the foundations of mathematics of all *logical* dependence on geometrical and kinematical intuition. It follows that if we wish to understand *how* geometry has disappeared from the logical foundations of mathematics, we must understand the logical underpinnings of axiomatic definition. To understand those underpinnings is to understand how set theory provides the foundations for all mathematics.

Here we come to the central reason why modern mathematics especially stands in need of a careful examination and exposition of its foundations. For there is widespread confusion concerning the very nature of the modern axiomatic method and, in particular, concerning the essential and ineliminable role set theory plays in that method[3]. I shall discuss this critical issue later in some detail[4]. But for now, suffice it to say that the *logical* dependence of axiomatics on the set-theoretical concept of mathematical structure requires that set theory already be in place before an account of the axiomatic method, understood in the modern sense of axiomatic *definition*, can be given. It follows necessarily, therefore, that *we cannot use the modern axiomatic method to establish the theory of sets.* We cannot, in particular, simply employ the machinery of modern logic, modern *mathematical* logic, in establishing the theory of sets.

There is, to be sure, such a thing as "axiomatic set theory"; but although this theory is of central importance for the study of the foundations of mathematics, *it is a matter of logic* that it cannot itself, as an axiomatic theory in the modern sense, serve as a foundation for mathematics. Set theory, as a foundational theory, is, indeed, an axiomatic

---

The novelty introduced by later mathematicians was to base the algebra of real numbers on set theory, using the technique of axiomatic definition.

[3] I have discussed this matter at some length in my article "What is required of a foundation for mathematics?" to which I refer the interested reader.

[4] Chapter 6, especially Sections 6.2, 6.3 and 6.4.

theory, but in the original sense of "axiomatic" that applies to traditional Euclidean geometry as traditionally understood. The axioms of set theory are not conditions that single out a class of interpretations, as are, for example, Hilbert's axioms for geometry. On the contrary, they are fundamental truths expressed in a language whose fundamental vocabulary must be understood *prior* to the laying down of the axioms. That, in any case, must be the view taken of those axioms by anyone who embarks on the enterprise of expounding the set-theoretical foundations of mathematics. Whether, or to what extent, any such enterprise is successful, whether, or to what extent, the axioms can legitimately be regarded in this manner, is, of course, a matter for judgement. But it will be a central part of my task to show that they can be so regarded.

1.2 What the foundations of mathematics consist in

As I have just explained, the foundations of mathematics comprise those ideas, principles, and techniques that make rigorous proof and rigorous definition possible. To expound those foundations systematically, one must provide three things: an account of the *elements* of mathematics, an account of its *principles*, and an account of its *methods*.

The *elements* of mathematics are its basic notions: the fundamental *concepts* of mathematics, the *objects* that fall under those concepts, and the fundamental *relations* and *operations* that apply to them. These basic notions are those that neither require, nor admit of, proper mathematical definition, but in terms of which all other mathematical notions are ultimately defined. Insofar as these basic notions of mathematics are clear and unambiguous, the basic propositions of mathematics, which employ them, will also be clear and unambiguous. In particular, those propositions will have objectively determined truth values: the truth or falsity of such a proposition will be a question of objective fact, not a mere matter of convention or of agreement among experts.

The *principles* of mathematics are its *axioms*, properly so called. They are fundamental propositions that, although true, neither require, nor admit of, proof; and they constitute the ultimate and primary assumptions upon which all mathematical argument finally rests. There is no sense in which the axioms can be construed as giving or determining the meaning of the vocabulary in which they are couched. On the contrary, the meanings of the various items of vocabulary must be given, in advance of the laying down of the axioms, in terms of the elements of the theory, antecedently understood.

The *methods* of mathematics are to be given by laying down the canons of definition and of argument that govern the introduction of new concepts and the construction of proofs. This amounts to specifying the *logic of mathematics*, which we must take care to distinguish from *mathematical logic*: mathematical logic is a particular branch of mathematics, whereas the logic of mathematics governs all mathematical reasoning, including reasoning about the formal languages of mathematical logic and their interpretations. The logic of mathematics cannot be purely formal, since the propositions to which it applies have fixed meanings and the proofs it sanctions are meaningful arguments, not just formal assemblages of signs.

Here it must be said that the need to include an explicit account of logical method is a peculiarity of modern mathematics. Under the Euclidean dispensation, before the advent of set theory as a foundational theory, and when definition played a much more modest role in mathematics, one could, or, in any event, one did, take one's logic more for granted. But with the rise of modern mathematics, in which definition has moved to the centre of the stage, and where mathematicians have gone beyond even Euclid in their quest for accuracy and rigour, it has become necessary to include logical methods among the foundations of the subject. In fact, the central problem here is to explain the logical principles that underlie the modern axiomatic method. This will raise questions of the logic of generality, of the *global* logic of mathematics, that are especially important, and especially delicate, as we shall see[5].

A systematic presentation of the foundations of mathematics thus consists in a presentation of its elements, its principles, and its logical methods. In presenting these things we must strive for *simplicity, clarity, brevity*, and *unity*. These are not mere empty slogans. The requirements for *simplicity* and *clarity* mean, for example, that we cannot take sophisticated mathematical concepts, such as the concept of a category or the concept of a topos, as *foundational* concepts, and that we cannot incorporate "deep" and controversial philosophical theories in our mathematical foundations. Otherwise no one will understand our definitions and no one will be convinced by our proofs.

The ideal of *brevity*, surely, speaks for itself. *Unity* has always been a central goal: unity in principles, unity in logical technique, unity in standards of rigour. With the stupendous expansion that has taken place in mathematics since the middle of the nineteenth century the

[5] I shall discuss this point in Sections 3.4 and 3.5.

need to strive for unity in foundations is even more pressing than ever: mathematics must not be allowed to degenerate into a motley of mutually incomprehensible subdisciplines.

This, then, is what an exposition of the foundations of mathematics must contain, and these are the ideals that must inform such an exposition. But the task of *expounding* the foundations of mathematics must be kept separate from the task of *justifying* them: this is required by the logical role that those foundations are called upon to play. A little reflection will disclose, indeed it is obvious, that there can be no question of a *rigorous* justification of proposed foundations: if such a justification were given, then the elements, principles, and logical methods presupposed by that justification would themselves become the foundations of mathematics, properly so called.

Thus the clarity of basic concepts (if they really are basic) and the truth of first principles (if they really are *first* principles) cannot be established by rigorous argument of the sort that mathematicians are accustomed to. Insofar as these things are evident they must be *self*-evident. But that is not to say they are beyond justification; it is only to say that the justification must proceed by persuasion rather than by demonstration: it must be dialectical rather than apodeictic.

In any case, self-evidence, unlike truth, admits of degrees, and, as we shall see, the set-theoretical axioms that sustain modern mathematics are self-evident in differing degrees. One of them – indeed, the most important of them, namely Cantor's Axiom, the so-called Axiom of Infinity – has scarcely any claim to self-evidence at all, and it is one of my principal aims to investigate the possibility, and the consequences, of rejecting it. But what is essential here is this: when we lay down a proposition *as an axiom* what we are thereby claiming directly is that it is *true*; the claim that it is self-evident is, at most, only implicit, and, in any case, is *logically* irrelevant.

1.3 What the foundations of mathematics need not include

It is obvious to anyone who teaches mathematics that means must be devised for presenting its foundations simply, yet rigorously and thoroughly, to apprentice mathematicians: they must be told about sets, about ordered pairs and Cartesian products, about functions and relations; they must be made to grasp the idea of mathematical structure, and of a morphology-preserving map between such structures; more generally, they must be taught the techniques of rigorous proof and rigorous definition, and, especially, must be led to understand the ideas and strategies that inform the method of *axiomatic definition* – the central technical idea underlying modern mathematics. Much of this is confusing, none of it is easy, and all of it is necessary: those who do not master these foundations will find the road to modern mathematics barred to them.

These practical necessities remind us that in laying down the foundations of mathematics we are actually engaged in mathematics proper. Those foundations are an integral and essential part of mathematics itself. Of course, when we reflect deeply on such fundamental matters, we are bound to encounter profound questions of a general philosophical character. Sometimes we may be forced to face up to them. *But we should make every effort to avoid incorporating purely speculative philosophical ideas into mathematical foundations, properly so called.* That this is necessary from the standpoint of mathematics should be obvious. Mathematicians, like infantrymen, must march off to battle carrying only such equipment as is absolutely essential to their task. But philosophers, too, will benefit if the foundations of mathematics are kept free of philosophical controversy insofar as that is possible. For it is useful to them to know what are the minimal philosophical presuppositions upon which mathematics can rest.

Put in these general terms, all this may seem rather obvious and unexceptionable. But the matter may take on an entirely different colour when I draw what I see as the necessary consequences of these observations. In particular, I take the view that the foundations of mathematics do not require, and therefore should not include, a general theory of the meaning of mathematical propositions, or a general theory of mathematical truth, or a general theory of how mathematical knowledge is acquired. In mathematics it is sufficient if our propositions *have* clear meanings; it is not our business, as mathematicians, to account for what having a clear meaning consists in. Our theorems must be true and our proofs valid; but we are not required to *say* what a proposition's being true or an argument's being valid amounts to. Mathematicians must strive to acquire mathematical knowledge; but they do not need a theory of what the acquisition of such knowledge consists in merely in order to acquire it. Such a theory would belong to psychology, not to mathematics.

The mathematician studies mathematical structures, such as groups or topological spaces, just as the entomologist studies insects or the palaeontologist fossils. It would be an absurd impertinence to demand of an entomologist that he supplement his descriptions of the behaviour and physiology of insects with an account of how it is that human beings can acquire knowledge of this sort, or communicate it to one another once they have acquired it. I say that the same should apply to the mathematician: we may insist that his definitions be precise, his theorems be true, and his proofs be valid. But that is all we can sensibly require, or have any reason to expect.

Of course some may argue that, unlike insects or fossils, mathematical structures are "abstract" or "ideal" entities, that they exist, if at all, only in the minds of mathematicians, and that, in consequence of having these gossamer and insubstantial things for its subject matter, mathematics gives rise to ontological and epistemological difficulties unprecedented in the other sciences, difficulties which must be addressed when the foundations of mathematical science are laid down. I am convinced, however, that this is a mistake. It is perfectly true that in the past mathematics was thought to have certain characteristic abstract or ideal "mathematical objects" for its subject matter. But such views are outmoded, for mathematicians can now use the modern axiomatic method to replace reference to those peculiar "objects" with discourse about mathematical structures. Mathematical *structures*, however, are not the "abstract" and "ideal" entities that the mathematical *objects* of tradition were thought to be, and do not give rise to the ontological and epistemological difficulties inherent in that tradition[6].

The great philosophical questions of meaning, truth and knowledge are no doubt of considerable interest in themselves; but it is not necessary to solve them before getting on with the business of proving, say, that a continuous function on a closed interval assumes a maximum, or that every integer is uniquely factorable into a product of primes. This is indeed fortunate, since *definitive* answers to these philosophical questions are nowhere in sight. Certainly there is not the remotest prospect of *universal agreement* on such answers. But there must be universal, or near universal agreement on what constitutes a valid proof or definition in mathematics – and, indeed, there is. If there were not, the subject would be in chaos.

We must also keep separate from foundations those general questions that, though not really philosophical, are what mathematicians themselves

---

[6] This is an important, difficult, and, I must confess, on the face of it, *controversial* point. What is essential to the claim I am making is the distinction between the "abstract" or "ideal" character of traditional mathematical *objects* and the arithmetical (set-theoretical) character of modern mathematical *structures*. Here I have simply stated, without argument, what I take to be the case. The argument is given, and at some considerable length, in what follows, principally in Chapters 2 and 6.

might call "philosophical": questions of the significance (or otherwise) of theories, of the suitability of mathematical definitions, strategic questions about the importance of problems, or about the most useful ways to tackle them, questions about the overall organisation of mathematics, questions about the relative importance of its various branches, ... These are questions about which all mathematicians, even those with no real philosophical interests, are called upon to think from time to time. We must not underestimate their importance, for it is often decisive, though usually only the very best mathematicians make significant contributions here. But these questions, which call for sound judgement and large experience, cannot be taken to be part of the foundations of mathematics, properly so called, although they are inextricably bound up with its practice. What belongs to mathematics proper – and that includes its foundations – cannot be speculative, or evaluative, or controversial. Indeed, the very word "mathematics" comes from the Greek *"mathêma"* which means simply "what can be taught and learnt", in other words "what is cut-and-dried".

In mathematics our aim is to start from what is simple and obvious, our basic concepts and axioms, and to proceed by obvious steps, our definitions and our inferences, to obtain what is often complex and difficult, our general concepts and our theorems. If this sort of thing is to work, we must strive to make both the starting points and the individual steps as transparent and as obvious as we can make them.

Accordingly, it is no reproach to an account of the foundations of mathematics that its basic concepts and its axioms are remote from the actual practice, and the immediate concerns, of most mathematicians. On the contrary, that very remoteness is rather a measure of the logical depth of our definitions and theorems, and, as such, is probably the best indication we have that our basic concepts and axioms are, in fact, suitable. Such a reproach has frequently been levelled at modern set-theoretical foundations. But that is to misconceive the purpose they are called upon to serve. Whatever the shortcomings of those foundations, remoteness from practice is *not* among them.

To be sure, the basic concepts and axioms of set theory are, indeed, remote from practice. One cannot gain insight into group theory or functional analysis or algebraic geometry by contemplating them. But that fact, though incontrovertible, is utterly irrelevant. The *only* question relevant here is whether those concepts and axioms do, in fact, logically sustain such disciplines. And *that* they unquestionably do.

1.4 Platonism

On the face of it mathematics is full of references to special mathematical objects, "abstract" or "ideal" things that we cannot touch or see. The mathematician's triangles, for example, are not to be identified with those he draws on the blackboard, or with the architect's or the land surveyor's. But even mathematical triangles seem relatively "concrete" when compared to other things that mathematicians regularly talk about: natural numbers, real numbers, functions, "spaces" and "structures" of various kinds .... What are these things? Do they really exist? And if so, how, and in what sense? These questions are as old as mathematics itself. Moreover, they have been a central preoccupation of philosophers in the European tradition since before the time of Plato.

Indeed, Plato himself has been invoked in the present day debate on these matters, for it is now the fashion to describe as "Platonism" the naive idea that the peculiar objects mathematicians talk about exist in their own special way – that they are what they are, so to speak – and that they really have the properties and relations that mathematicians say they have. "Platonism", understood in this sense, is often used as a term of abuse or, alternatively, adopted as a badge of defiance.

But the relation of modern "Platonism" to the opinions on mathematics actually held by Plato and his disciples is not at all what it is commonly supposed to be. It is true that Plato posited a special category of eternal and unchanging objects, the "Mathematicals" or "Intermediates", occupying a place in the realm of being midway between ordinary objects of sense and the Platonic Ideas or Forms. But it is clear that Plato did not regard what we should call "sets" and he called "numbers" (*arithmoi*) as necessarily belonging to the class of Intermediates. On the contrary, only sets of a certain special kind were classed as mathematical objects in his special sense. This is a matter of some significance, as I shall make clear when I come to discuss the question of set existence in the next chapter.

In its modern usage the term "Platonism" does less than justice to the historical facts. Moreover, that usage rests upon a classification of objects into the "abstract" and the "concrete" that is so crude, so simple-minded, and so undiscriminating as to be useless. It is best abandoned. In any case, as I shall show in the next chapter, modern set theory, which in the present day estimation is the very quintessence of "Platonism", is fundamentally Aristotelian, not Platonic, in spirit.

Nevertheless, one of Plato's principal doctrines in the philosophy of mathematics is an essential component of modern set-theoretical foundations. I mean his anti-operationalism:

> ... no one who has even a slight acquaintance with geometry will deny that the nature of this science is in flat contradiction with the absurd language used by mathematicians, for want of better terms. They constantly talk of "operations" like "squaring", "applying", "adding", and so on, as if the object were to *do* something, whereas the true purpose of the whole subject is knowledge – knowledge, moreover, of what eternally exists, not of anything that comes to be this or that at some time and ceases to be[7].

Plato seems to be invoking his doctrine of Intermediates here, but we should ask whether his doing so is really necessary to his central point. Given that the truths of mathematics are timeless truths, does it then follow that they must, of necessity, be truths about timeless entities like Plato's Intermediates? Such a supposition is natural enough, I suppose, but is it really essential? The question is a deep and difficult one, and deserves a more than merely cursory examination. But that question aside, Plato is surely right in holding that mathematics is not primarily a matter of *doing*, but rather of *knowing*.

I take operationalism in mathematics to be the doctrine that the foundations of mathematics are to be discovered in the activities (actual or idealised) of mathematicians when they count, calculate, write down proofs, invent symbols, draw diagrams, and so on. No doubt we ought to be chary of following Plato in positing "mathematical objects", and, indeed, modern mathematics provides us with the conceptual tools which make this possible; but we ought all to account ourselves "Platonists" in this sense: *considerations of human activities and capacities, actual or idealised, have no place in the foundations of mathematics, and we must therefore make every effort to exclude them from the elements, principles, and methods, upon which we intend to base our mathematics.*

This is no easy matter, for the art of mathematics consists, in large part, in finding suitable symbolic expression for our concepts and propositions with a view to replacing complicated conceptual thought with mere symbolic manipulation – letting our notation do our thinking for us, so to speak[8].

---

7 *Republic* 527a.

8 No clearer illustration of this can be given than by contrasting the ancient theory of ratio and proportion given in Book V of the *Elements* with the modern, symbolic handling of cognate material in the algebra of real numbers. Euclid's treatment is complicated and cumbrous, and is carried out purely conceptually, so that the exposition is almost entirely verbal. The modern theory, by contrast, is entirely algebraic, that is to say, is largely a matter of manipulating symbols, so that complicated arguments in Euclid

When we are engaged in mathematics our attention is constantly shifting between the notation we employ and the subject matter. Our concern is now with the symbols themselves as syntactico-combinatorial objects, now with the things for which they are the signs; now those symbols are the objects of investigation, now the medium of expression. Often it proves necessary to find suitable *objective correlatives* for our symbols, that is to say, non-linguistic, non-symbolic objects corresponding to certain symbols or symbol combinations, or to processes of algebraic or numerical calculation. In this way we render our discourse objective by purging it of its "human, all too human" elements; and it is the great strength of modern mathematics that it provides us with powerful techniques for accomplishing this necessary purge.

The need to exclude operationalism, in all of its guises, from the foundations of mathematics is not something that can be established in a few paragraphs of argument: it is the central lesson of the whole modern movement in mathematics, a lesson which mathematicians absorb almost unconsciously in learning their trade, and practise without even reflecting on it. It is built into the conventions of expository style that every mathematician must master. But many of the most widely used and fundamental concepts in mathematics have an operationalist air about them in consequence of their origins in the contingencies of mathematical practice as that practice has developed historically: the concepts of "natural number", "ordered pair", "function", and "relation" are all of this character. To objectify these concepts, so to speak, is, inevitably, to introduce some appearance of arbitrariness and artificiality into our mathematical discourse: we must face up to this as we cannot avoid it. But the bedrock, the concepts in terms of which these concepts are defined, must be free of any operationalist taint.

---

correspond in the modern theory to simple syntactic operations. Of course there is a price to pay for such facility, and whereas it is clear *ab initio* just what Euclid is talking about, mathematicians had to wait until the end of the nineteenth century before an adequate account of the facts that justify the symbolic manipulations of real algebra was given.

2

Simple Arithmetic

2.1 The origin of the natural numbers

The natural numbers $0, 1, 2, \ldots$, as we now understand them, are not simply given to us as part of the "raw data" of mathematics. On the contrary, these numbers were invented, indeed invented fairly recently, along with rational, irrational and negative numbers. There is, in fact, something distinctly unnatural about our "natural" numbers.

This is so important a matter that I want to make doubly sure that no one misunderstands me: when I say that these numbers were invented I am making a particular, historical point, not a general, philosophical one. It is not my intention to resurrect the philosophical claim that mathematics is invention not discovery (surely it is both), nor the more particular claim that the natural numbers are "mental constructions" or anything of that sort. On the contrary, what I am talking about is an actual, historical process of invention that began sometime in the late middle ages and culminated in the late seventeenth century, by which time mathematicians had arrived at what is essentially our modern conception of real number. In the course of this process, the concept of number was drastically altered – no, that is not strong enough: in the course of this process the word "number" was stripped of its customary and traditional meaning to be assigned an entirely new meaning, one which had scarcely anything in common with the original[1].

Not only the fact, but also something of the actual nature of the change in the meaning of "number", can be deduced from the definition given by Isaac Newton in his *Universal Arithmetic*:

By a *Number* we understand not so much a Multitude of Unities, as the abstracted

[1] Naturally I do not mean to suggest that this change of meaning was a phenomenon confined to English alone. A similar change occurred in other European languages.

Ratio of any Quantity to another Quantity of the same kind, which we take for Unity.

From this brief passage we can glean several important facts. It is clear that Newton recognised "quantities" of various kinds, and these quantities were not "numbers" in Newton's sense; for the latter are said to be abstracted *ratios*[2] of two quantities of the same kind. Moreover, the notion of "number" that he mentions only to reject ("By a *Number* we mean not so much" – that is to say, not at all – "a Multitude of Unities but ... ") must, by the logic of the sentence, have been the notion of number that some of his readers might have expected or, at least, have been aware of. In fact, it recalls the definition given by Euclid in Book VII of the *Elements*:

> A number (*arithmos*) is a multitude composed of units.

So Newton actually tells us that he doesn't mean by "number" what Euclid meant[3]. What Newton means by "number" is what we should mean by "real number", or, at least, very like what we should mean: for we must not lose sight of the fact that it was important for Newton, as it would not be for a modern mathematician, to ground his theory in the ancient science of quantity, the theory of ratio and proportion expounded in Books V and VI of the *Elements*. Newton's "Quantity" is Euclid's "*megethos*"; such a quantity is, for example, a line, or a surface, or a solid, or a time.

These things may have been regarded as *idealisations* of physical lines, surfaces, solids, and durations, but are clearly not *abstractions* in the sense that Newton's "numbers", or indeed ours, are abstractions. It was, of course, this ancient science of quantity that the new science of (real) number replaced.

How crucially important it is in mathematics to choose the right terminology! The use of "number" for this new concept was especially unfortunate; and there were perfectly good alternatives ready to hand: "(abstract) ratio" and "(abstract) quantity" immediately suggest them-

[2] That is to say, relationships in respect of size. (See Euclid's *Elements*, Book V, Definition 3.)

[3] In fact there may already have been a shift in meaning from Euclid's "multitude *composed* of units", which can only mean what we should call a *set* of units, to Newton's "Multitude of Unities" which may refer, not to a set of units, but to its (abstract) cardinality. (See Klein's *Greek Mathematical Thought and the Origin of Algebra*, Chapter 12, especially pp. 201–202). The question of how and why the new notion of number arose is a fascinating one, but for our purposes here the fact, and the nature, of the change in the meaning of "number" are all that are directly relevant.

selves. By choosing this oldest of mathematical words to name what was then the newest of mathematical concepts, the mathematicians of the seventeenth century virtually ensured that their successors would eventually lose sight of the very concept of number as it was understood in antiquity.

In *Frege: Philosophy of Mathematics*, for example, Michael Dummett writes

> That the number of objects of a given kind is the set of those objects is sufficiently absurd to need no refuting. (p. 82)

No doubt this view *is* absurd as an account of *our* notion of number. But this "absurd" view was held by Plato, Aristotle, Euclid, Aquinas, and Ockham, and was, as we have seen, acknowledged by Newton, who, however, failed to remark upon its "absurdity", even while he was in the course of explicitly rejecting it.

In fact the ancient concept of number provides, as I intend to show, a simpler, more straightforward, and more natural account of the facts that underlie simple arithmetic than does the modern notion of "natural" number; and if one adopts it one is not burdened like Frege, and, indeed, Dummett, with the task of explaining what *things* those "natural" numbers are.

The original notion of number is so important and so fundamental that it could not remain suppressed. It had eventually to reappear, even if only under another name: what our ancestors knew as "numbers" we now call "sets".

2.2 The abstractness of the natural numbers

I think it unlikely that any modern mathematician would be drawn to Newton's account of number. Of course that account could not be taken as a *definition* of "number", since it does not meet the modern requirements of rigour. But that apart, it is not so much its vagueness as its particularity that seems unsatisfactory. When Newton speaks of "abstracted ratios of quantities" he has something too definite in mind, something quite alien to the modern mathematical sensibility. His abstractions are somehow too concrete for our taste, if I may put it in that somewhat paradoxical way.

This comes of Newton's desire, which I alluded to earlier, to base the new science of number on the old science of concrete quantities. We moderns are, in any case, chary of mixing our natural numbers quite so thoroughly with our reals; and we want all our numbers kept logically independent of geometry.

It seems, then, that Newton's account of the abstractness of numbers won't do. And yet we all agree, do we not, that the modern notion of number, in general, and of natural number, in particular, is highly abstract? The question I want to address now is: "abstract" in what sense?

The "abstractness" of our modern natural numbers is something much simpler, much more insubstantial, than the abstractness of Newton's numbers. Indeed, the abstractness of our numbers is a *fact* about the way we view them, not a *theory* about their natures. It manifests itself in the naive idea that number words and numerals are names, or signs, for particular objects. This idea imposes itself on us, *inter alia*, by our use of certain familiar expressions (e.g. "the number five") and by the way that we understand simple numerical equations (e.g. $128 + 279 = 407$).

The interpretation of "$128 + 279 = 407$" that perhaps most naturally suggests itself is this: if we perform the operation of addition on the natural numbers 128 and 279 (in that order) then we obtain the natural number 407. This way of understanding such equations is, I submit, suggested to us both by the syntactic form of the equations themselves, and by our rules and methods for calculating sums. For the abstract operation of addition here corresponds to the actual procedure of calculation (hence "operation", with its suggestion that something is to be *done*), and the abstract numbers to which that operation applies correspond to the numerals employed in such calculations.

Natural numbers thus present themselves to us as those things, whatever they are, that correspond to the numerals and letters we use in symbolic calculation. They are generated by our notation, and by the syntactic and algorithmic rules that govern its employment. This, no doubt, accounts for their peculiarly thin and insubstantial character, even as "abstract objects".

Abstractions of this *symbol generated* sort, though unknown to the Greeks, are quite common in modern mathematics. Some of them play indispensable technical roles: ordered pairs are an obvious example[4]. But wherever such symbol generated abstractions occur, they are a potential source of perplexity and confusion. For it is never obvious that

[4] Functions and relations are also symbol generated abstractions of this sort. I shall discuss the logical status of these key notions in Chapter 4.

there really is anything to which they naturally correspond, outside our symbols themselves that is. They are epiphenomena of our notation.

The naive idea of the natural numbers that I have described here – the idea that they are the particular abstract things named by our number words and numerals – this idea scarcely constitutes a theory, although it is sometimes rhetorically inflated into one. It is really rather a starting point for theories, philosophical or mathematical: it is what those theories have to explain or to explain away. It poses the following dilemma: if there are, in fact, objects of which our number words and numerals are the names, what are those objects? If there are no such objects, what is arithmetic about?

2.3 The original conception of number

Let us consider the idea of number that our modern idea of "natural" number has supplanted, the classical Greek concept of *arithmos*[5]. On that conception, a number (*arithmos*) is a finite plurality (multitude, multiplicity) composed of units, where a unit is whatever counts(!) as one thing in the number under consideration. Thus Trigger, Champion, and Red Rum constitute a number of horses, and each unit in this number is a horse; red, yellow, blue, and green constitute a number of colours, and each unit is a colour. This original meaning of "number" still survives in English, as when we say, "Lieutenant Lightoller was included among the *number of survivors* in the wreck of the Titanic".

In the two examples I have given the units are homogeneous: all of them are horses or all of them are colours. Such numbers provide the most straightforward and unproblematic examples of numbers understood in this ancient and original sense. When the units are all of the same kind, then what it is to be a particular kind of number, for example, a triple or a quadruple, of the kind of thing they are, is determined by what it is to be one thing of that kind, and by what it is to be, say, three things of any kind whatsoever. If you know what a horse is, and you know what a triple is, then you know what a triple of horses is; and if you know which particular horses Trigger, Champion, and Red Rum are, then you know which particular number of horses they compose.

Of course, no one has to know this particular number of horses, or even the horses that make it up, in order for it to *be* a number of horses.

[5] See Jacob Klein's *Greek Mathematical Thought and the Origin of Algebra*, especially Chapter 6, Paul Pritchard's *Plato's Philosophy of Mathematics*, Chapters 1–3, and Myles Burnyeat's "Plato on why mathematics is good for the soul".

Simply by being, severally and individually, the particular horses that they are, and by being, collectively, finite (in fact, three) in multitude, Trigger, Champion, and Red Rum make up the particular number of horses that they do. They do not have to be collected together, either in reality or in conception, in order to compose that number: there is nothing that anyone has to think or to do in order to bring it into being. Indeed, this is obvious, on reflection, for a herd of twenty-five horses contains two thousand three hundred such horse triples, most of which, of course, no one would ever separate out, or even think of, not even someone well acquainted with the horses, both individually and as a herd. A number of horses is no more a creature of the mind than are the individual horses that compose it. Since we can count such numbers, it is natural that we "count" them as things.

Getting this point right is important for everything that follows. In the example just considered, one might be tempted to say that what are being counted are not numbers – *arithmoi* – of horses, but, for example, the possible ways of selecting three horses from a herd of twenty-five. That is, indeed, a common way of speaking about such matters, and, moreover, it has a reassuringly "concrete" air about it: one can easily imagine cowboys cutting horses out of herds. But such imaginings are irrelevant, and such confidence in the "concrete", understood in this sense, is misplaced. For it is the existence of the *arithmoi* – the triples – that grounds the possibilities of selection, and not the possibilities of selection that ground the existence of the *arithmoi*. It is impossible to think those triples away: they are simply "there" to be counted as units in the *arithmos* of 2,300 horse triples that they compose.

But to what extent must the units of a number be homogeneous? Are we allowed to count disparate, even incongruous things as together constituting a number? Indeed, are we not forced to acknowledge numbers composed of heterogeneous units? Are they not simply "there" by virtue of their units being "there" in finite multitude? Frege notes, with approval, that

> Leibniz rejects the view of the schoolmen that number is not applicable to immaterial things, and calls number a sort of immaterial figure, which results from the union of things of any sorts whatsoever, for example, of God, an angel, a man, and motion, which together are four[6].

If we were to follow Leibniz and Frege and allow the widest possible latitude in the choice of units, then we should have to acknowledge

---

[6] *Die Grundlagen der Arithmetik*, p.31.

that finitely many things (e.g. two hundred and ninety-seven) of any kinds whatsoever, however various and heterogeneous, simply by being, individually and severally, the particular and definite things that they are – horses, or men, or ideas, or characters in fiction, or numbers (in the sense discussed here) – and by being, collectively, finitely many things (in the circumstances posited, two hundred and ninety-seven), are the units of a unique number that together they all constitute.

But what kinds of things are suitable to serve as units in a number? Surely, some kinds of thing are too vague, or too indistinct, or too poorly differentiated to count as units. Clouds, ripples on the surface of a liquid, psychological states – such things are usually too indefinite to count. How many psychological states did you experience yesterday? How many clouds are there now overhead in the sky? It's not that these questions have answers that we don't know; it's rather that, in general, they don't have answers – objectively determined answers – at all.

But sometimes we can, in fact, count such things. There are occasions on which we can, for example, say that there are three clouds overhead. And, after all, do we not speak of four Noble Truths, seven types of ambiguity, three theological virtues, thirteen ways of looking at a blackbird, . . . ? What are the numbers that these sorts of things compose? Do Faith, Hope, and Charity form a triple in the way that Trigger, Champion, and Red Rum do? That is rather like the question whether Faith is a *thing* in the way that Trigger or Champion is.

The bafflement and uncertainty we experience when we confront such questions remind us that the ancient conception of number under consideration was not an exact and artificial scientific concept but a concept in common use. Natural concepts in ordinary use characteristically exhibit a fluidity and suppleness that makes them unsuitable for exact, scientific discourse in their raw state, so to speak. The domain of applicability of such a concept typically is sharply and clearly delineated at its centre, but fades into vagueness at its periphery. In the case we are considering, the vagueness that infects the notion of *number* at its boundary is the same vagueness that infects the notion of *thing*.

It is a characteristic of language that it allows us to form substantives by combining expressions in complex ways, and to use them as if they were ordinary nouns in forming sentences. When we form a sentence in this way it seems as though we were predicating something of a thing. In this way we pepper our discourse with references to "possibilities", "ways", "likelihoods", "facts", "circumstances", and so on. Thus arises the illusion (if illusion it be) that there are "things" corresponding to,