Philosophy of mathematics

Selected readings

SECOND EDITION

Edited by
Paul Benacerraf
STUART PROFESSOR OF PHILOSOPHY
PRINCETON UNIVERSITY

Hilary Putnam
WALTER BEVERLY PEARSON PROFESSOR OF
MODERN MATHEMATICS AND MATHEMATICAL LOGIC
HARVARD UNIVERSITY

Cambridge University Press
Cambridge
London New York New Rochelle
Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published by Prentice-Hall, Inc., 1964
Second edition published by Cambridge University Press 1983

Printed in the United States of America

Library of Congress Cataloging in Publication Data
Main entry under title:
Philosophy of mathematics.
Bibliography: p.
1. Mathematics – Philosophy. I. Benacerraf, Paul.
II. Putnam, Hilary.
QA8.4.P48 1983 510'.1 82-25257
ISBN 0 521 22796 8 hard covers
ISBN 0 521 29648 X paperback

Contents

Preface to the second edition

Even a casual comparison of the table of contents of the present collection with that of its predecessor will reveal significant differences as well as much overlap. By and large, the present selection is the product of two forces: (a) comments from users of the first edition (and from potential users of the second) and (b) our own sense of the direction the field has taken during the past two decades.

We are grateful to our many friends and colleagues, too numerous to thank individually, who have commented on what they found useful and less than useful in our first effort, as well as on what they felt it would be good to have available in one volume. Their perspective has been invaluable, though the responsibility for our selections remains largely our own.

Needless to say, we would have liked in a way to reissue the first edition and simply add a second, companion, volume. But we are deterred by the prohibitive cost (to the user) of the two volumes. Hence the inevitable compromise: A selection was made, omitting several things to make room for new ones. In a number of cases (most notably the Wittgenstein material and "Two Dogmas of Empiricism"), the (present) availability of most of the material enabled us to omit it with less of a sense of loss. Not so with the rest. The selection of new material was even more difficult, as these years have been particularly fecund, both in relevant semi-technical results and in philosophical explorations.

As before, we limited our selections to those we felt would be accessible to the philosophically educated reader with enough background in logic to understand an exposition of some of the results of twentieth-century logic. (An important example is the independence of Cantor's Continuum Hypothesis.) In a similar vein, we tried also to narrow the range of *philosophical* issues discussed in the selection to ones that could most easily be recognized as concerning the philosophy of *mathematics*. Both of these admittedly loose principles served as guidelines only; but any attempt to observe them inevitably constrains the range of literature available for consideration. Except for these rules of thumb, in the end, we followed no overarching principle other than that of making a selection of items that, in our judgment, would make interesting reading when taken together.

Another major point of difference between this volume and its predecessor lies in the extended Bibliography that appears at the end of this edition. It was compiled by starting with the selected bibliography of the first edition, adding a number of items that we felt were missing from it, and closing it by including everything referred to in any of the selections included in the book. Inevitably, there have been many important omissions – as we go to press we must resist the urge to keep adding to it. We wish at this point to record our gratitude: once again to George Boolos, who compiled the initial bibliography; to Takashi Yagasawa, for his help in augmenting it; and to Ann Getson, James Cappio, and Ann Ivins for their invaluable contributions to its completion and to the preparation of the manuscript.

Finally, a word about format. Collecting all of the bibliographies into one comprehensive Bibliography enabled us to put all bibliographical references into a standard form for the book as a whole. This has meant the elimination of a large number of purely referential footnotes (for the convenience of including them in the text), as well as the shortening of many others. Also, we have identified as such any references to items that appear in this collection, inserting the present page number(s) where relevant. Wherever feasible, we retained the author's reference to the original source and simply added the locus of that reference in the present reprinting. We hope that the effort that went into this task is rewarded with a book that is, as a consequence, much more useful.

Princeton, N.J., and Cambridge, Mass.
November 1982

PAUL BENACERRAF
HILARY PUTNAM

Introduction

1. General remarks

It would be difficult to say just what comprises the philosophy of mathematics – what questions, views, and general areas should be covered in a book such as this. With that as our excuse, we have not tried to bring together a collection of selections that could be said to cover the field in a comprehensive way. We have tried rather to bring together selections that we felt were interesting in their own right, and that offered interesting comparisons when read together, all with the proviso that the issues discussed in them were in most cases central to the field. If we have succeeded, then we are certain that the reader has an adequate introduction to the philosophy of mathematics.

The divisions we have chosen are largely arbitrary, as the following remarks will indicate, and no importance should be attached to a particular article's being in this section rather than that. With this much said, we can state that the motivation behind the sections is roughly as follows: We included in Part I those selections that centered around three traditionally important views on the nature of mathematics: logicism, intuitionism, and formalism. This is not to say that other articles in the book do not bear on these views. For example, the article by Hempel in Part III is itself a very clear exposition of logicism. Like the other pieces in Part III, however, it discusses the view on a (mathematically) less technical plane and is therefore more readily accessible to people with no formal training in logic than the papers in Part I. The discussion of the three aforementioned views is thus the unifying thread that runs through the section called "The Foundations of Mathematics."

Questions concerning mathematical existence (we leave open here the question of whether mathematical existence is a different sort of existence or the existence of a different sort of thing, or both, or neither) are touched upon principally in Part II. But is is clear to anyone with the slightest familiarity with these matters that intuitionism, at least, is a view concerning mathematical existence, at least insofar as it includes conditions on what is to count as *proof* of the existence of certain mathematical structures, entities, and so on. Therefore, an adequate consideration of these questions (adequate in that it takes in the leading points of view) would include items from Part I as well.

The first three sections overlap further in that the third, "Mathematical Truth," contains, besides the Hempel selection just mentioned, the quasi-intuitionistic piece by Poincaré. Quine ("Truth by Convention") discusses conventionalism, a view also expressed by Carnap in connection with mathematical existence and truth in his article in Part II, and Benacerraf's article in Part II, "What Numbers Could Not Be," is a discussion of an issue central to logicism: the nature of numbers.

We feel that such overlap is unavoidable. The division into problems is, at best, a guide for the reader. It is evident that one's view on the nature of mathematical truth (if there is indeed such a beast) will affect one's views on mathematical existence and will constitute a position on the "Foundations of Mathematics."

Despite this overlap, there is a further division to which we can point and which may prove helpful in organizing the array of views represented in this book: There is a suggestive distinction to be drawn between the items in Part I on the one hand, and those in Parts II and III on the other (Part IV, as we shall see, thoroughly straddles this distinction). Part I contains contributions belonging to what we should like to call the "epistemology of mathematics." With the possible exception of the selections by Frege and Russell, the authors of these pieces devote a good part of their attention to the question of what an acceptable mathematics *should* be like: what methods, practices, proofs, and so on, are *legitimate* and therefore justifiably used. They don't take existing mathematics and mathematical activity as sacrosanct and immune from criticism; according to them, there are justifiable and unjustifiable methods in mathematics, and acceptable results are those obtainable by justifiable methods. In fact, a good portion of the effort of the mathematician should be devoted to trying to recast intuitively desirable and acceptable results in forms that show them to be ultimately acceptable. If the author in question is an intuitionist, then it will be his view, e.g., that any part of real analysis that cannot be obtained by intuitionistic methods ought to be discarded. But in most cases, it is vitally important to carry on the search for intuitionistic proofs of as yet unobtained classical theorems.

And what we have seen to be true of the intuitionist is also true of formalists (Hilbert, von Neumann, Curry). Members of this latter group are concerned about the legitimacy of references to infinite collections, structures, and the like, in mathematics. More particularly, the concern takes the form of a concern that such reference, since it is to things so far from what we are capable of experiencing, might lead to contradictions (presumably because the candle of intuition casts but a dim light from

such a distance). What then becomes important for the formalist is the search for a proof that these "infinitistic" methods form a consistent whole. The proof must, of course, be one that does not employ these questioned methods. So here again there appears a restriction on what methods should be countenanced in mathematics. And so on for the others. The other feature characterizing the members of this group is that they are predominantly mathematicians rather than philosophers. And we say this without in any way wishing to play down their philosophical contributions, any more than we would be inclined to deny the mathematical contributions made by members of the other group. The first group consists, therefore, largely of mathematicians who criticize the very foundations of their subject. These are the "epistemologists." (Poincaré and Bernays really belong in group one, if we are to judge from the bulk of their work, but the passages we have chosen assimilate them more to the second; Dummett, on the other hand, although primarily a philosopher, is writing to supply philosophical underpinnings to the intuitionist position and, as such, fits more naturally in group one.)

In contrast with the "epistemologists of mathematics," there are those who accept mathematics as, if not sacrosanct, then at least not their province to criticize. Their task is a different one: They do not want to *promulgate* certain mathematical methods as *acceptable*; they want to *describe* the *accepted* ones. Mathematics is something given and to be accounted for, explained, and accurately described. For them, epistemology is not a tool to help sort the good mathematics from the bad – it is a scheme within which mathematics as such must fit ("Mathematical propositions are analytic," "Mathematical statements are true by convention," and so on). One way of describing the difference between these two groups is to say that, for one group, the epistemological principles have a higher priority or centrality than most particular bits of mathematics, and hence can be used as a critical tool; whereas for the other group just the reverse is the case: Existing mathematics is used as a touchstone for the formation of an epistemology, one of whose conditions of adequacy will be its ability to put all of mathematics in the proper perspective. To put it somewhat crudely, if some piece of mathematics doesn't fit the scheme, then a writer in the first group will tend to throw out the mathematics, whereas one in the second will tend to throw out the scheme.

Of course, matters are not quite that neat. For both groups there is a constant interplay between epistemological principle and mathematical activity. Members of the first group will sometimes start with some paradigm cases of acceptable mathematical practice (e.g., intuitionists and formalists both start with parts of number theory) and then try to arrive at principles that will account for the validity of this starting point. These principles are then used either to criticize what fails to conform to them, or as a guide for the erection of standards that further proofs must meet, especially proofs of those "theorems" already "proved" but not "acceptably" so. Similarly, it would be an exaggeration to saddle the second group with absolutely everything a mathematician might produce. Their account of mathematics might very well force them to renounce and denounce some piece of mathematics as unacceptable. But by and large this is very unlikely. Ayer, Hempel, Boolos, Benacerraf, and Putnam are all prepared to take mathematics pretty much as is. Russell and the Carnap of "The Logicist Foundations of Mathematics" present a problem, depending on how seriously one takes their discussions of the vicious circle principle and impredicative definitions. There is no question that the Carnap of "Empiricism, Semantics, and Ontology" has abandoned any *critical* function for epistemology. Quine is a borderline case, almost professionally so. Insofar as he has abandoned his ontic qualms, however, he presents no problem. But leaving the borderline cases straddling the border line, Heyting, Brouwer, Hilbert, von Neumann, Curry, and Poincaré (in other passages than those we are presenting, alas) are quite clearly on the other side. Kreisel belongs, if anywhere, to this latter group, for he has strong constructivistic leanings. The "if anywhere" is inserted because Kreisel's function has been more or less to reconstruct and make mathematical sense of the philosophical pronouncements of the members of this latter group. And it is quite conceivable that he should wish to see just how much of extant mathematics one can obtain on this or that restriction without much caring whether or not one adopts the restriction. This would make his, as it were, a metatask.

The articles in Part IV weave both strands together into a hopeless tangle. This is hardly surprising, as the subject matter is set theory – a branch of mathematics with powerful ancestral ties to philosophy, and which has served as the battleground for a host of philosophic disputes ever since its explosive Cantorian(-Fregean) birth and its stormy Russellian adolescence. It has survived these traumas, as well as those inflicted by Gödel and Paul Cohen through their discovery of the independence results. Now, philosophers and mathematicians alike are scrutinizing its history and prehistory to tease out both the strengths responsible for its survival and the genetic weaknesses that almost caused it to perish. Inevitably, the reformers and apologists rub elbows.

But the distinction is a vague one and we should not try to make too much of it. Though vague, we hope it is nonetheless suggestive; and it

should be of some help in understanding how the writings of the authors included in this collection are related. What becomes of interest, once one has seen the distinction, is the way in which one can view the discussions included in this book as continuous with one another. At first sight, it might appear that the two groups did not even discuss the same problems. But it should be seen that Hilbert is just as much concerned with the determinants of "mathematical truth" as, say, Ayer. The positions they adopt are rather different, but both should be read as writing on the same questions, or very nearly so. And similarly with Gödel. When Gödel discusses the continuum hypothesis, he is merely focusing on a particular mathematical proposition and the ways in which it might be shown to be true or false. Hempel's remarks on the nature of mathematical truth should, if cogent, be relevant to this discussion. And so on with the rest of the selections.

Consequently, it is our view that the questions we have chosen for study are intimately related. Furthermore, the authors whose discussions of these questions we have selected are, on the whole, concerned with very much the same questions, superficial differences to the contrary. These differences bespeak differences in point of view and differences in methods of attack, rather than simply different concerns. We believe that the discussion of all these problems benefits greatly from the interplay of these differences – but only when it becomes clear what unites them as discussions of the same problems. It is our hope that reading these selections together will make this clear.

So much by way of introduction. The remainder of this Introduction will consist of remarks on a number of problems, some of which are discussed rather fully, and others of which are merely touched upon by our authors. We hope that these remarks will make it easier to understand the selections and the issues involved.

A word of warning is in order. We have not attempted, in the sections that follow, to present a single, unified point of view on all of the problems and authors that we discuss. Instead we speak from many, different, and often incompatible viewpoints. For a couple of reasons. First, we are more concerned here to *raise* useful and interesting questions than to attempt to answer them. Second, it is unlikely that there is one on which we could agree, or even agree to agree for the nonce. Thus, the attentive reader, moving through this Introduction from section to section and even within individual sections, will discern shifts of focus and point of view. We hope that, rather than distract, they will prove helpful through the variety of perspectives they offer on the extremely complex problems under discussion.

2. The actual infinite and formalism

Although this collection does not contain a section titled "The Infinite in Mathematics," anyone who reads the selections from the writings of Brouwer, Heyting, and Hilbert that we have included under the general title "The Foundations of Mathematics" will quickly realize that the role played by infinite structures, collections, quantities, and the like in classical mathematics has a great deal to do with the controversy between the different "schools" in the philosophy of mathematics. By the same token, the measure of success attained by Cauchy and Weierstrass in eliminating "infinite quantities" from the calculus had a great deal to do with the ideal, shared by thinkers with views as mutually antagonistic and those of Hilbert and Brouwer, of eliminating the infinite from mathematics altogether.

But why should it be deemed desirable to avoid reference to the infinite in mathematics? Sometimes it is said – even by Hilbert – that references to the infinite are "meaningless." But why should one suppose that this is so? Classical philosophers – in particular Hume – had argued against the notion (in connection with infinite divisibility) on the basis of an identification of what is intelligible with what can be *visualized*; but the "image-in-the-head" theory of meaning no longer seems tenable, and attacks on the notion of the infinite must depend on something more reasonable than this if they are to be taken seriously.

In point of fact, it is very hard to find reasoned and even moderately detailed argument on this point. Opponents of the "actual infinite" tend to assume that the burden of proof lies on the other side. "Show us that the notion makes sense," they seem to say, where the criterion of making sense seems to be expressibility in *their* terms. We cannot discuss this issue here: Suffice it to say that readers who sympathize with the demand of the classical empiricists that all concepts be legitimized by being "derived from experience" will probably find themselves inclined to sympathize with those who doubt that any notion of an infinite structure is a clear one, whereas readers who are either of a more realistic or of a more pragmatist turn of mind may have difficulty in seeing "what the fuss is all about."

Suppose, however, that we assume that statements about infinite structures "make sense." Are there in fact any such structures to talk about? Hilbert argues convincingly that physics provides no clear evidence for the existence of such structures: In fact, the progress of physics has, as he points out, introduced finiteness and discontinuity in area after area in which the infinite and the continuous once reigned supreme. Today even the possibility of a beginning (and an end) to "physical time"

is under discussion among physicists. Thus we must agree with Hilbert that if mathematics is to be independent of dubious empirical assumptions, it must not base assertions concerning the existence of infinite structures on physical considerations.

To this Russell replied, in a slightly different context, that mathematics is concerned not with (physical) existence, but only with the *possibility* of existence. Thus, in the second edition of *Principia Mathematica* (henceforth: *PM*), Russell and Whitehead chose not to assume the so-called Axiom of Infinity, which asserts that there are infinitely many objects in the universe of discourse, but rather explicitly included it among the hypotheses of each "theorem" in whose proof it was used. If T was the "theorem" in question, then Whitehead and Russell asserted only '*if* Inf. Ax., *then* T'.

But is it clear that an infinite totality could *possibly* exist? If the Axiom of Infinity leads to a contradiction, then the theorems that list it as hypothesis are certainly not very interesting. Since these form a large part of mathematics (at least as reconstructed by Russell and Whitehead), should there not be some proof of the *consistency* of the Axiom of Infinity, whether it is to be used as a postulate of the system or only as a hypothesis of a large number of important theorems?

Here we get a parting of the ways in the philosophy of mathematics. Russell and his followers apparently regard the *possible*, if not actual, existence of infinitely many objects as self-evident, whereas for Hilbert and the formalists the consistency of this assumption must be *proved*. Moreover it must be proved by "finitist" means – that is to say, the assumption itself must obviously not figure, even in a disguised way, among the assumptions of the consistency proof. The reader will observe that this kind of question is a bit like a political question – it is not a "purely theoretical" question, in the sense of making no difference to practice, but rather it affects one's standards in mathematics and one's program as a mathematician. Hilbert did not think it very likely that the system of *PM* was, in fact, inconsistent; he simply felt that to take its consistency, or even the consistency of elementary number theory, without proof, was to adopt too low a standard of mathematical exactness and to risk unpleasant surprises in the future.

Another perspective that might prove helpful in understanding Hilbert's desire for a consistency proof for infinitistic classical mathematics is the following.

Hilbert took certain kinds of mathematical assertions to be philosophically (i.e., epistemologically) unproblematic. These were assertions whose truth or falsity could be determined by combinatorial calculation – by the observation of combinatorial facts that could be ascertained by immediate perception. Call them "basic propositions" – for example, whether one (finite) string of symbols was longer than another. Assertions that made essential reference to infinite collections, in ways that deprived them of this property of finite verifiability/falsifiability in terms of observable combinatorial facts, were considered to be strictly meaningless. Hilbert recognized that their introduction into mathematics considerably simplifies the statement of a number of laws, thus making the theory considerably more elegant and appealing. In this regard he likened it to the introduction of ideal elements, such as points at infinity in projective geometry, or *i*.

But is it safe?

Hilbert allowed that they might nonetheless be admitted into mathematics if it could be shown that their admission would be harmless – that it would not enable one to prove falsehoods, that is, false basic propositions. A consistency proof for classical mathematics that made no essential reference to infinite collections would do just that. If the proof made use only of finitistic (i.e., sanitary) principles, the reference to infinite collections would have been justified on finitistic grounds. It could then be regarded as simply an elaborate *façon de parler* and indulged without risk.

3. The "potential infinite" and intuitionism

For the intuitionists, the position with respect to the infinite was different. Given a set of statements describing an infinite structure, there are two sorts of doubts that may arise. First, one may question the *consistency* of the statements: This was Hilbert's worry. Second, one may doubt that the statements "pick out" a *unique* and well-defined mathematical structure. Intuitionists sometimes write as if even the notion of an "arbitrary finite magnitude" is not completely fixed in advance.[1] We know, indeed, that 1, 2, 3 are integers. We know that certain operations applied to integers lead to integers – e.g., addition, multiplication, exponentiation. But it does not follow that we have a perfectly definite notion of "any integer" – because this involves the idea of iterating an operation (say, "adding 1") an *arbitrary finite number of times*, and we need not admit that we have a clear notion of what this means. The intuitionist does not, of course, propose to do without the concept "integer" – that would be to abandon mathematics altogether. The proposal (cf. Heyting,

[1] In what follows, we present an account of intuitionism directed at the nonintuitionist. It is possibly an account acceptable to no intuitionist, but we feel that such "falsification" is justified if it helps bridge the gulf that presently exists between the intuitionist and the "classical mathematician."

elaborating and formalizing Brouwer's ideas) is rather to develop a propositional calculus for dealing with concepts that do not necessarily correspond to a well-defined totality (and "statements" that do not necessarily have a truth-value). This attitude is often described as "countenancing the potential infinite but not the actual infinite." What it comes to is this:

1. A statement about an infinite structure – say, an infinite sequence of zeros and ones – may be regarded as *true* if proved and *false* if refuted, but in all other cases it is regarded as *neither true nor false*.
2. Since the structure is not thought of as well-defined, a statement about it can be proved only *if it is actually proved for a much larger class of structures*. In fact, to prove a statement about an infinite structure, we must prove the statement on the basis of verifiable statements either about some *finite part* of the structure (e.g., the first ten places of the sequence), or about the *rule* (if there is one) for successively producing the finite initial segments of the structure.

An example may help to make this position more clear. Consider the assertion that the sum of the first n odd number $(1+3+\cdots+(2n-1))$ is always a perfect square (in fact n^2). The sum of the first *one* odd numbers, that is, 1, is a perfect square, since $1=1^2$. And if the sum of the first n odd numbers is n^2, then the sum of the first $n+1$ must be $(n+1)^2$, or n^2+2n+1 (since the $n+1$st odd number is $(2n-1)+2$, and this is equal to $2n+1$). Thus we have proved the theorem for 1, and if we have proved the theorem for n, we can prove it for $n+1$. Accordingly, the intuitionist – like the classical mathematician – concludes that the theorem holds for every number. The philosophical difference is that the intuitionist does not assume that the numbers are a well-defined totality. But in this case it doesn't matter. (Although there are many cases in which intuitionists are led by their position to reject classically valid proofs; for example, proofs that assume that every statement about an infinite totality is either true or false – which amounts to assuming that the totality is well-defined – are rejected by intuitionists.) For, even if we extend the notion of an integer to cover a new "object," if all theorems proved in the preceding fashion hold for all the things previously counted as integers (notice that there need be only finitely many of these at any given time), and if the "new" integer is always "one plus" something previously counted as an integer, then the theorems in question will hold also for the "new" integer.

By way of contrast, consider the assertion that the number of "twin primes"[2] is infinite. For the classical mathematician this has a unique truth-value (even if he doesn't presently know it). But for the intuitionist it doesn't: For he has no proof of the statement, nor does he have a proof of its negation, and statements about an ill-defined totality don't have a truth-value unless they are proved (or disproved) from a *partial* determination of the totality.

A classical mathematician can get an approximate idea of what the intuitionist has in mind in the following way: (1) *Drop* the assumption that there is a well-defined "standard model" for number theory. (2) Don't assume that we can characterize by any finite number of axioms all of the things that we would intuitively recognize as correct methods of proof. (I.e., take "number theory" itself as a concept in the process of being created.) Then there will be three classes of statements in number theory: statements that are "true," that is, true in all models of number theory; statements that are "false," that is, false in all models; and statements that are "neither true nor false," that is, true in some models and false in others. Also, the "true" statements will all be provable – but not necessarily in any one formal system.

This explanation is, however, not intuitionistic, since an intuitionist would not accept the idea that the undecidable statements are "true in some models and false in others." Moreover, the intuitionist surely objects with reason here: for if "number theory" is not a closed concept, then the notion of a "model" of number theory is surely not a mathematically meaningful one even from the standpoint of classical mathematics. Yet one can still make sense of proving that a statement is "true in *all* models," namely, if one can prove that a statement is true in all models for some finite fragment of number theory, then, however the concept of "number theory" may be *enlarged* in the future, the statement in question must be, indeed, "true in all models of number theory" (since models of the whole must be models of each part). However, the idea of an infinite "model" – a *well-defined* infinite collection satisfying the axioms of a formal system – is not one acceptable to an intuitionist.

For the intuitionist, also, the problem of consistency does not arise because any statement about a "potential infinite" can be interpreted as a statement about a *finite* (but extendable) structure.[3] Thus the intuitionist and the formalist are in agreement as to the part of mathematics that is "safe"; that is, whose consistency may be taken as evident on the basis of an interpretation; namely, the part that may be interpreted as referring only to finite structures.[4]

[2] A *prime number* is one that cannot be divided without remainder except by itself and 1. *Twin primes* are primes whose difference is 2: e.g., (5, 7), (11, 13), (17, 19), etc. Whether the number of such pairs is finite or infinite is an unsolved (and apparently hopelessly difficult) problem in number theory.

[3] Strictly speaking, this is true only of free-variable statements. And even there, although the assumptions used in any one free-variable proof always have a finite model, the proof that *this* is so may require non-finitist methods. This is one problem with "finitist mathematics": the consistency of "finitist" systems is not, in general, demonstrable by strictly finitist means.

[4] We are indebted to Georg Kreisel for the remark that the intuitionist notion of the "potential infinite" has *two* classical analogues: the "ill-defined infinite set" and the "finite (but unbounded) segment."

4. Logicism

Logicism (Frege–Russell–Whitehead) arose out of a concern with a different problem: the nature of mathematical truth. Logicists hoped to show, as against Kant, that mathematics did not have any "subject matter," but dealt with pure relations among concepts,[5] and that these relations were "analytic," that is, of the same character as the principle of noncontradiction, or the rule of *modus ponens.* In contrast, Hilbert maintained that mathematics *did* have an extralogical subject matter, namely *expressions*[6] (e.g., series of strokes |, ||, |||, ···) and that its simplest truths (e.g., "|| added to ||| is |||||") were *anschaulich* (a German word that can mean both "visual" – in colloquial German – and "self-evident" or "intuitive" in philosophical German).

Logicism had one great and undeniable achievement – it succeeded in reducing all of classical mathematics (by any reasonable standard excluding completeness) to a single formal system. This achievement was much admired by the formalists, even if they did not agree that "mathematics has been reduced to logic." Formalists held that, as a result of the work of Whitehead and Russell, one had at last a clear formalization of what it was that had to be proved consistent.

Logicists, of course, thought they had done more than just axiomatize

[5] For Frege, this is a complicated issue. Concepts for him are not objects or "entities" of any sort (what this pronouncement *means* is far from clear, and the subject of much controversy among Frege scholars). So, for logic (and hence mathematics) to deal with the relations among concepts is not for them to have a special subject matter – in the way, say, that living organisms constitute the subject matter of biology. Unfortunately, the issue is further complicated by Frege's view that concepts have *extensions,* which *are* objects (indeed, numbers, for Frege, are the extensions of concepts). Thus, to deny logic and mathematics any special subject matter, the logicist must argue that extensions of concepts (what some would simply call "sets") do not themselves constitute a special domain – a special subject matter.

[6] In place of "expressions" one might, of course, use other things: e.g., tables and chairs, or musical tunes. The important thing for Hilbert was not that finitist mathematics should be literally about series of *marks* (e.g., |, ||, |||, etc.) but that the subject matter, whatever it might be, should be wholly finite, discriminable, and *anschaulich* in all of its relevant parts and relations.

extant mathematics. They believed that they had *derived* all of mathematics from pure logic, without using any extralogical assumptions, and thus shown it all to be analytic. To assess this claim, we must ask at least whether what the logicist derives in his formal system is the mathematics he sets out to derive, and whether the premises of the derivation belong to logic.

Whether the logicist reduction should count as a derivation of mathematics depends, of course, on the character of the definitions employed – more specifically on what those definitions *preserve*. If they preserve meaning, at least sentence by sentence, then the answer is clearly yes. For he has shown that sentences with the same meaning as those of mathematics are logical consequences of the axioms of his formal system. To the extent that something less than meaning is preserved, the claim that it is the *propositions of mathematics* that have been derived must at least be questioned.

The issue of whether the derivation is from *logical* premises is regarded by many as largely a verbal issue (at least it need not be settled to see the bearing of the logicist reduction on Kant's claim that arithmetic was synthetic a priori): Logicists did not reduce all of mathematics to *elementary* logic, but they did reduce mathematics to elementary logic *plus* the theory of properties (or sets), properties of properties, properties of properties of properties, and so on. Thus if property theory (or set theory) may be counted as part of logic, mathematics is reducible to logic. But to what extent this refutes Kant's claim and establishes mathematics as analytic is something still open to question, for several objections may be raised. (It is not our purpose here to argue these points in detail, but we feel that it is particularly important to raise objections to logicism because it is a view that has received very little criticism in the literature: Of all the authors we reprint, only the intuitionists are really seriously critical of it. But they attack it from a very different point of view, as we will point out later.)

What Kant had denied was that the propositions of mathematics (arithmetic would be the relevant branch here, since he might have conceded the analyticity of algebra) were analytic. But 'analytic' for him meant either 'following from the law of noncontradiction' or being (a "logical truth") of the form 'All A are B', where "the idea of being a B is contained in the idea of being an A." A relevant example of such an analytic truth might be 'All spinsters are females'. Now, it is hardly the case that the logicist reduction of mathematics clearly shows the propositions of mathematics to be of either of these kinds. "Following from the law of noncontradiction" is itself at best a very unclear notion. The most likely (and most charitable) construal for it is something like 'whose

negations are self-contradictory'. Thus construed, the question becomes that of deciding whether showing that on one set of plausible definitions arithmetic can be derived from set theory establishes that the negations of (presumably the true) arithmetic propositions are self-contradictory. It establishes, to be sure, that *if* these definitions are correct analyses of the meanings of the arithmetic terms, and *if* the set-theoretic axioms are themselves analytic in the relevant sense, and *if* being derivable in first-order logic from analytic propositions via definitions representing correct analyses constitutes "following from the law of noncontradiction," *then* indeed the logicists have shown that the propositions of arithmetic follow from the law of noncontradiction. But these are very big *if*s. Probably the biggest of them is the one concerning the analyticity of the set-theoretic axioms. In what sense would *they* be "analytic"? Many, even of those who don't doubt their consistency, would balk at their analyticity. But even should this *if* be granted, the other two loom large. The reader should refer to Quine's "Carnap and Logical Truth" for some objections to the first.

And, of course, it is just not the case that mathematical propositions have been shown to be analytic in the second of the two Kantian senses cited (i.e., reducible to "logical truths" that have the form 'All A are B'). It might be objected that to defend Kant on this basis is to trivialize him in the process, because he surely would have widened his notion of what constitutes logic if presented with, say, quantification theory. Therefore, the claim that ought to be examined is whether mathematics is reducible to quantification theory. To this, two replies might be offered. The first is, of course, that mathematics is not so reducible. Some set theory or its equivalent is needed as well. Hence this widening would not suffice. The second might be that Kant would not have agreed to a widening of the notion of logic beyond the *monadic* predicate calculus, so that the question does not even arise. And in either case, the problem of what constitutes a correct analysis of the meaning of a mathematical term is still with us and likely to remain for a while to come.

Yet it should not be forgotten that if today it seems somewhat arbitrary just where one draws the line between logic and mathematics, this is itself a victory for Frege, Russell, and Whitehead: Before their work, the gulf between the two subjects seemed absolute.

One difficulty with calling set theory "logic" concerns the axioms of set (or property) existence; e.g., $(\exists P)(x)(\sim P(x))$ (read: 'There is a P such that for all x, x does not have P' or more simply 'there is an empty property [or set]'). In his last years, Frege came to the conclusion that such assertions of existence were not part of *logic* at all and repudiated "logicism," which he had founded. Another difficulty is the need for an Axiom of Infinity in deriving mathematics: In order to meet this difficulty, Frege, having given up logicism, proposed to derive mathematics from *geometry* (where the Axiom of Infinity is true, since presumably there are infinitely many points) instead of from "logic."

Russell, as has already been mentioned, proposed in the second edition of *PM* not to take the Axiom of Infinity as a postulate of the system, but to list it as a hypothesis whenever it was needed to prove a theorem. But then it becomes puzzling how mathematics is *useful* (if a great many of its theorems have the form 'if Inf. Ax., then p', and "Inf. Ax." – the Axiom of Infinity – is, in fact, empirically false).

In connection with the first difficulty, it has been argued that '$(\exists P)(x)(\sim P(x))$' is a necessary truth, since there is indeed a proposition '$P(x)$' that is false for every x, namely '$x \neq x$' (or any other self-contradictory proposition). More generally, Russell has sometimes suggested that '$(\exists P)$' need not be interpreted as meaning that some extra-logical entity "exists," but may only be a way of indicating that there is a meaningful proposition '$P(x)$' with the specified characteristics. (Hilbert would reply: You still need the notion of the *existence* of *formulas*, i.e., *expressions*.) Sometimes Russell writes in this way: as if a property (or, as he says, "propositional function") were only a linguistic expression containing free variables (e.g., 'x', 'y',...) – or, perhaps, the meaning of such an expression. However, this interpretation of *PM* is, in fact, excluded if "impredicative definitions" are permitted. (For explanation, see the article by Carnap in Part I.) For, if 'P' ranges only over "properties nameable by formulas of *PM*" then this restriction – to objects nameable in *PM* – will appear in the definition of every set. In particular, 'real number' will only be able to mean 'real number nameable in *PM*'. However, under the intended interpretation of the version of *PM* that permits impredicative definitions, there is an expression that stands for the set of *all* real numbers, not just *nameable* real numbers.

Strangely, Russell never appreciated this difficulty, and called *PM* a "no-class theory" to the end, although his "propositional functions" are nothing but arbitrary sets (or "classes") under another name, if impredicative definitions are permitted.

Another achievement of Frege and Russell was the analysis of the concept "number." Since this analysis is presented in detail in several of our selections, we shall not review it here. However, it raises several points of disagreement between logicists and intuitionists.

According to the intuitionists, one cannot understand 'two', 'three', etc., unless one has the general notion of a *number*. On the other hand,

the logicists maintain that 'two', for example, is (contextually) definable thus:

$$2(P) \equiv (\exists x)(\exists y)[P(x) \cdot P(y) \cdot x \neq y \cdot (z)(P(z) \supset . z = x \vee z = y)]$$

(read: 'there are two Ps if and only if there are x, y such that x is P and y is P and x is not the same thing as y and for all z, if z is P then z is the same as x or z is the same as y'.)

Here the intuitionists may perhaps be right. The logicist account, however, could easily be modified so as to take care of this criticism: namely, define "number" just as the logicists do[7] (roughly, a "number" is anything that can be obtained from zero – or the class of all empty classes – by repeatedly applying a certain "successor" operation), and then define 'zero' *not* as 'the class of all empty classes' but rather as 'the smallest *number*,' (defining 'smallest' in some suitable way, or as 'the *number* that is not a successor'), 'one' as 'the *number* that is the successor of zero', 'two' as 'the *number* that is the successor of one', etc. Then the notion 'number' will be part of the definition of *each* number.[8] Of course, the definition of 'two' will be *equivalent* to the one Russell employed, but not synonymous with it word-by-word (or rather symbol-by-symbol). So one who used the new definition could perfectly well agree with the intuitionists that the Russell definition does not express the customary meaning exactly.

Another disagreement between logicists and intuitionists is over the identification of numbers with sets of sets (e.g., of zero with the set of all empty sets). As a point separating these two camps, it has less importance than is customarily accorded to it. In the first place, it is questionable whether Frege[9] held that "zero," "one," "two," etc., *had* to be identified with any particular entities: the important thing was the analysis of 'there are two Ps', 'there are three Ps', etc. The intuitionists accept this analysis as mathematically correct. Perhaps the intuitionist would prefer to render 'there are two Ps' by 'the species of Ps can be put in one-to-one correspondence with the numbers one, two' – however, Frege would certainly have accepted this definition.

[7]We mean to suggest here that an intuitionist could accept the logicist definition of number in terms of the "ancestral" (a number is something that is either 0 or bears the ancestral of the successor relation to 0); not that the famous Frege–Russell definition of the ancestral would in turn be acceptable to an intuitionist.

[8]This procedure may sound circular, but clearly it is not, provided that the expression 'zero' does not appear in the definition of 'number'; i.e., that a number be defined, e.g., as either the set of all empty sets or something bearing the ancestral of the successor relation to the set of all empty sets. *Zero* could then be identified with the "smallest" number, etc.

[9]We are indebted to Michael Dummett for this and other points in connection with Frege.

A more important disagreement concerns the logicist claim that "mathematics can be reduced to logic." Intuitionists reject this claim on the following grounds:

1. Understanding any system of deduction involves already having the notion of *iterating an operation an arbitrary finite number of times*; and this the intuitionists regard as a fundamentally *mathematical* and not logical notion. (Recall that they also regard it as a "creative" or extendable notion – not one whose every application is completely clear and specifiable in advance.)
2. *The principle of mathematical induction* (which we used in our proof that n^2 is the sum of the first n odd numbers) is a fundamentally *mathematical* one (closely connected with the idea of the *iteration* of an operation), and not reducible to logic. Frege did indeed reduce mathematical induction to what *he* called logic – via the "definition of the ancestral" (see the Frege, Russell, and Hempel selections). This reduction, however, depends on the use of impredicative definitions, which are rejected by intuitionists, and also on the axioms of set existence, which would not be called "logic" by intuitionists even if they did accept them.

5. Tautologies and sets

In his *Tractatus Logico-Philosophicus,* Wittgenstein maintained, following Russell and Frege, that mathematics was reducible to logic. Logic, in turn, was reducible to propositional calculus, according to Wittgenstein. This is correct, for the system *PM,* whenever the number of individuals is a fixed finite number, but it is correct in the infinite case only if infinitely "big" expressions[10] are permitted. The idea is, briefly, to treat universal statements as infinite conjunctions: 'Everything is F' is treated as 'x_1 is F & x_2 is F & . . . & . . .' (where $x_1, x_2, \ldots$ are all the individuals in the universe of discourse, in some order). Now, the truths of the propositional calculus are all "tautologies" – they come out true, combinatorially, under all possible assignments of 'true' and 'false' to the "elementary propositions." Thus was born the very popular philosophical slogan that "mathematics consists wholly of tautologies."

Of course, closer examination revealed serious difficulties with the *Tractatus* view. A quantifier over properties (i.e., such an expression as 'for all properties P') is expanded as a conjunction with one clause for

[10]One cannot even say "infinitely long," because some of the expressions that would arise at higher types if all quantifiers were "expanded" as truth-functions would be *nondenumerably* infinite, and could not be thought of as existing ("written out") in primitive notation.

each property of individuals. But this presupposes not only that the individuals form a well-defined totality, but that the properties (or *sets* of individuals) form a well-defined totality, and similarly for sets of sets, sets of sets of sets, and so on. This is already debatable for 'property' in the sense of the term in which each property corresponds to a possible "rule for selecting"; for 'property' (or rather 'set') in the sense of *arbitrary collection* (*any* collection, whether given by a rule or by "chance"), the situation is even worse. Consider, for example, the famous "continuum problem" of Cantor. This asks whether there exists some set (in the sense of *arbitrary* set) of real numbers (*arbitrary* sequences of integers) that can be put into one-to-one correspondence with neither the set of all integers nor the set of *all* real numbers. The answer 'no' has been proved by Gödel to be *consistent* with the axioms of set theory. And the answer 'yes' was proved by Paul J. Cohen *also* to be consistent with those axioms. In what sense then would it be true that "there really is" (or "really isn't") any such set? One might answer: 'in the sense that if you listed all the sets of real numbers, you would (or wouldn't) find one such that if you listed all the one-to-one correspondences (arbitrary sets of pairs consisting of a real number and an integer, or of two real numbers, satisfying the "one-to-one" condition) you would not find a correspondence mapping the set in question onto the integers, and you would similarly fail to find a correspondence mapping the set in question onto the real numbers.' This answer, however, is completely unhelpful for many reasons: e.g., the notion of "listing" all the sets of real numbers is absurd if taken literally; so is the notion of completely examining even one nondenumerably infinite and "random" collection of real numbers in detail; and then how much more absurd is the notion of examining *all* "one-to-one correspondences"!

Today, very few philosophers or mathematicians of any school would maintain that the notion of, say, an arbitrary set of sets of real numbers is a completely clear one, or that all the mathematical statements one can write down in terms of this notion have a truth-value that is well-defined in the sense of being fixed by a rule – even a non-constructive rule – that does not assume that the notion of an "arbitrary set" has already been made clear. The contention that, even in the absence of such rules, questions such as the continuum problem have a definite meaning and, having a definite meaning, have a definite answer, quite independently of the state of our knowledge, forms the core of what has variously been called "realism" or "platonism" in the philosophy of mathematics. (For a remarkably lucid and forceful statement of this position, see Gödel's article in Part IV on the continuum problem, especially the supplementary section.)

Nevertheless, there is a respect in which this is the natural position to take. We normally do not require an effective method of verification as the sine qua non of meaningfulness. This was a requirement made in quite another context (empirical science) by the Vienna Circle, and long since abandoned by most of its proponents. Why should it be different with mathematics? If we think we understand what is meant by a set, a one-to-one correspondence, and so on, why shouldn't we say that the continuum problem has a definite answer, no matter how far we may be from finding out what it is? What do the two have to do with one another? A split on this question normally reveals a split on the most fundamental issues in the philosophy of mathematics, on the very nature of mathematical activity.

In general, the platonists will be those who consider mathematics to be the *discovery* of truths about structures that exist independently of the activity or thought of mathematicians. For others not so platonistically minded, mathematics is an activity in which the mathematician plays a more creative role. To put it crudely, propositions are true at best insofar as they follow from assumptions and definitions we have made. If we can show that a proposition is *undecidable* from the assumptions we currently accept, the question of its "truth" or "falsity" vanishes in a puff of metaphysical smoke. Our assumptions, definitions, and methods of proof constitute the rules determining the truth or falsity of the propositions formulated in their terms. If a proposition is undecidable from our current assumptions, then its "truth" is not determined by the available rules. Since nothing else is relevant, the question of truth does not arise. The platonist does not agree because, for him, the truth of mathematical propositions is not determined by the rules we adopt, but rather by the correspondence or noncorrespondence between the propositions and the mathematical structures to which the terms in those propositions refer. In his view, mathematical terms and propositions have meaning above and beyond that conferred on them by the assumptions and methods of proof accepted at any one time.

To get an idea of the objections that might be raised to the platonistic way of looking at the continuum problem let us look briefly at the notion of an "arbitrary set," which is needed for the formulation of the problem. The reader may perhaps wonder what is wrong with our preceding explanation: "Arbitrary set" means "any set, whether given by a rule or by chance." The difficulty is that the notion of *chance* makes no sense in pure mathematics, except as a figure of speech. Suppose, however, we took this explanation literally: We might, for example, define an "arbitrary sequence of integers" as a sequence that *could* be generated by a

"random device." One difficulty is then the word 'could'. 'Could' can only mean mathematical possibility here, since we do not want to let physical laws have any effect on mathematical truth. But "mathematical possibility" is itself a disputed notion, where infinite structures are concerned. And a further difficulty is that, according to classical mathematics, there are other infinite sets, for instance the set of all sets *of sets* of real numbers, which are so "big" that they cannot be put into one-to-one correspondence with the set of all integers or even with the set of all real numbers: Such sets could not be identical with the "output" of any possible physical process, even if we were to take the notion of a "possible (actually infinite) physical process" as itself a clear one.

Again, some people say: "Why worry about possible physical models at all? You know what a *collection* is (as in 'collection of oranges') and you know what an integer is; therefore you know what is meant by 'collection of integers', and by 'collection of collections of integers', etc." This "simple-minded" point of view hardly seems satisfying, however. In the first place, our ordinary notion of a "collection" is loaded with physical connotations. If we say that these are to be disregarded, and that the members of a "collection" need not be proximate in space and time, need not be "similar" in any particular respect, and so on, then we are left with the notion of something like a random listing of objects. And if we say that the members of a "collection" (a) need not be objects, numbers, and so on, but may themselves be "collections," and (b) need not even be capable of being listed (or for that matter, named in language), even by a random device working through an infinity of time, then what notion are we supposed to form at all?

Second, the presence of statements (such as the continuum hypothesis) corresponding to which there is no verification or refutation procedure (except looking for a proof – which is most certainly not going to do any good, if 'proof' means 'proof in present-day set theory') is, perhaps, a reason for at least suspecting an unclarity in our notion of a "set."

Here it is instructive to compare set theory with number theory. In number theory too there are statements that are neither provable nor refutable from the axioms of present-day mathematics. Intuitionists might agree that this shows (not by itself, of course, but together with other considerations) that we do not have a clear notion of "truth" in number theory, and that our notion of a "totality of all integers" is not precise. Most mathematicians would reject this conclusion. Yet most mathematicians feel that the notion of an "arbitrary set" is somewhat unclear. What is the reason for this difference in attitude?

Perhaps the reason is that a verification/refutation procedure is inconceivable for number theory only if we require that the procedure be effective.[11] If we take the stand that "nonconstructive" procedures – i.e., procedures that require us to perform infinitely many operations in a finite time – are conceivable,[12] though not *physically* possible (owing mainly to the existence of a limit to the velocity with which physical operations can be performed), then we can say that there does "in principle" exist a verification/refutation procedure for number theory. For instance, to "verify" that an equation $P(x,y,z)=0$ has a solution using the "procedure," we *check each ordered triple x, y, z of integers.* (Of course, this requires working forever, or else completing an infinite series of operations in a finite time.) Similarly, to check a statement of the form $(x)(\exists y)P(x,y)=0$ (read: 'For every x there is a y such that $P(x,y)=0$') by the "procedure," we have to substitute 0 for x, and then check through $y=0,1,2,\ldots$ until we find a y_0 such that $P(0,y_0)=0$; then we substitute 1 for x, and look for a y_1 such that $P(1,y_1)=0$; and so on (again this requires an infinite series of operations). What this shows is: The notion of "truth" in number theory is not a dubious one if the notion of a completed actually infinite series (of, say, definitely specifiable physical operations) is itself not dubious. Since many mathematicians do not share intuitionist doubts about the clarity of the actual infinite, it is understandable that such mathematicians are willing to take the notion of number-theoretic truth as precise. For instance, Carnap argued for the legitimacy of such "nonconstructive rules" in explaining the notion of number-theoretic truth in his famous book *The Logical Syntax of Language.*

By way of contrast, we recall that *no* physical structure (not even an infinite one) can serve as a "standard model" for set theory. In addition, even if we allow "working forever," "completing infinite series in a finite time," and so on, no precisely definable sequence of operations exists by means of which we could "in principle" tell (in the sense explained in connection with number theory) whether an arbitrary statement of set theory was true or false by the procedure of exhaustively checking all cases. In short, if you understand such notions as *counting, adding, multiplying,* and *seeing if two numbers are equal,* we can explain to you the notion of a "true statement of number theory" (though not if you consistently "boggle" at all quantifications over an infinite domain); but to have explained to you the notion of a "true statement of set theory" or of an "arbitrary set," it would appear that you must already have some such notion in your conceptual vocabulary.

[11]An effective procedure is (roughly) one that a computing machine could be "programed" to employ. Gödel proved the impossibility of an effective decision procedure for number theory.

[12]E.g., if one has an infinite series of operations to perform, say $S_1, S_2, S_3, \ldots$ and one is able to perform S_1 in 1 minute, S_2 in 1/2 minute, S_3 in 1/4 minute, etc.; then in 2 minutes one will have completed the whole infinite series.

6. Mathematical truth

An admittedly naive view, which has certain attractions (perhaps its very innocence is among the greatest of them), runs like this:

Some propositions are true. Others aren't. Generally, a proposition is true if the reality it purports to describe is as that proposition depicts it – if the things referred to (if any) have the properties they are said to have or stand in the relations in which they are said to stand. All of this is quite independent of whether we know or have any reason to believe (or in the extreme – *could* ever have any reason to believe) that they are true. Our language can express (and we can understand) a whole range of questions, quite independently of whether we possess their answers, or ever will possess their answers, or even ever *could* possess their answers. Consider. On January 1, 1901, at 12 noon GMT, every molecule on Earth had an approximate location (never mind the rest of the universe). We don't now know that distribution. We never will know it. Perhaps there are reasons why in principle we never *could* know it. (Perhaps any representation of such a map is too big for us to comprehend, and comprehend we must if we are to know.) Still it seems plausible that such a distribution existed. We can frame the question. It has an answer. But the answer is too complicated.

And so too with certain mathematical questions.

Until recently it was not known whether every map can be colored with four colors, with no two regions that share a common boundary being colored the same color. Now we know. Mathematical research proceeds, at least in part, by answering questions previously put. At each stage there are meaningful questions that haven't been answered – and perhaps others that it is beyond us to answer despite the fact that it was we who framed them. In brief, every "well-formed" sentence in the language of science and mathematics expresses a meaningful proposition (one having a truth-value) about its subject matter. Whether a question is meaningful, whether it has an answer, does not depend on whether we know the answer, ever will know the answer, or even whether it is "in principle" possible for us to find it out.

Semantics is independent of epistemology.

Such an approach fits nicely with a conception of man and his place in nature in which man, like other animals, is a limited being. There are bounds on our epistemic powers – bounds dictated by the number of cells in our bodies, by the arrangement of those cells (the structure of our eyes is a good model: Things far too small for us to see, wave lengths too long to register, are not thereby deprived of existence). And so too there are things too complex for us to understand, questions just too hard for us to answer – though if we were differently constituted they might be within our reach. And might some of these questions not nevertheless be simple enough for us to frame? Any philosophy that adjusts the bounds of reality to those imposed by the existing (and in some sense necessary) limitations on our *epistemic* powers is bound to be shortchanging the world.

A referential semantics exhibits the propositions of physics as being "about" rigid bodies, fields, electrons; those of number theory as about numbers; set theory about sets. They are true if and only if the relevant entities have the properties ascribed to them.

In this form, our naive view is a kind of realism. Conjoin it with a platonist's perspective on the nature of the objects populating the domains of many mathematical theories (such as numbers, sets, functions, and spaces) – that these are abstract, exist outside of space-time, and independently of our conceptions. Plausible as such a perspective may seem (what and where *could* they be?), the result begins to threaten the pastoral calm of our opening scene. Because it now becomes unclear how we could know *any* of these mathematical propositions on their platonist construals. Our accounts of the truth-conditions of mathematical propositions and the nature of the objects that form their subject matter, independently plausible as these may have seemed, when joined clash with our most fundamental epistemological theories (for more on this general problem see Benacerraf "Mathematical Truth" and Section 8 of this Introduction).

Although not always explicitly stated, we feel that this is an important component of many of the questions that prompt the myriad of answers we call theories of the "foundations of mathematics," by mathematicians and philosophers alike:

- Formalists deny, among other things, the associated platonism and attempt to supply mathematics with a more visible subject matter and truth conditions for its propositions whose presence or absence it is at least sometimes clearly within our power to ascertain. Indeed, Hilbert wanted his program to produce, in addition to a proof of the consistency of mathematics, a general method which, applied to any given mathematical question (framable in a particular formal language), would either answer it or show it to be independent of existing assumptions. Church's proof of the unsolvability of the decision

problem for first-order logic dashed that hope, after Gödel had shown by his second incompleteness theorem that another principal aim of Hilbert's program – a finitist consistency proof for arithmetic – was already beyond reach.
- Nominalists traditionally have objected to the postulation of objects satisfying the platonist's description (although more recently the focus of the objection has shifted – see Part II, on the existence of mathematical objects, for more details), sometimes because they thought there simply weren't any such things, but more often for epistemological reasons – because they found the idea of an abstract object *unintelligible*.
- Conventionalists attempt to account for mathematical truth by bypassing the referential semantics (thus avoiding altogether the issues raised by platonism – mathematics has no "objects," or if it does, they simply have the properties we assign to them by convention). Mathematical truth thus reduces to the truth of certain conventions (by our fiat) plus the preservation of truth through logical consequence. Quine addresses these problems in "Truth by Convention" and in "Carnap and Logical Truth" far better than we could here. Although it is to some extent parasitic on Quine's reply, we should mention one reply to the conventionalist that is not often made but is worth considering:

  Everything "true by convention" is supposedly true. But conventions, however well-intentioned, can turn out to be inconsistent. First, their consistency or inconsistency is a mathematical fact of combinatorial mathematics (the fact that certain programs for computing do not lead to the "output" '1 = 0'), one that is itself hard to represent as a matter of convention. If this is right, then not *all* of mathematics can be true by convention. This suggests that our ability to make even axioms "true by convention" is already limited by the (nonconventional) fact of their logical consequences. But might we not make contradictions *themselves* true by convention? Few so far have been so devoted to conventionalism as to suggest that as a way out.

  And so on with a number of other views.
- Logicism is a special and difficult case – because it is really many cases – and we have dealt with it in Section 4 of this Introduction.

But there is another way in which epistemic considerations can entice one's account of mathematical truth away from the naive view described at the beginning of this section – other, that is, than the direct clash between epistemology and platonism. This is through the rejection of the realist principles proclaiming the independence of semantics (the theories of meaning and of reference and truth) from epistemology. It represents semantical features of sentences (meaning, truth conditions) as parasitic on epistemological ones (conditions of warranted assertability). For the naive realist we depicted earlier, it was the *absence* of such a coupling that opened the floodgates to propositions (mathematical or other) that were meaningful, had determinate truth-values (and thus represented in some respects possible states of "the world"), but yet remained entirely beyond our epistemic grasp. A philosopher can hold such an epistemically determined view about both mathematics and empirical science, or about either of them separately. Many phenomenalistically inclined positivists were verificationists about empirical matters but not about the formal sciences (perhaps because, in the idiom of the day, these were already thought to be "empty of factual content"). Intuitionists, on the other hand, appear to be verificationists in *mathematics,* but not necessarily elsewhere.

Michael Dummett, in his article "The Philosophical Basis of Intuitionistic Logic," urges a more thoroughgoing abandonment of the naive realist position.

To simplify considerably, Dummett's intuitionist construes mathematical propositions as having truth-conditions that *coincide* with their verification conditions, thus very neatly bridging the gulf that, for the platonist, separates truth and knowledge. This is proposed in support of the claim that the canons of reasoning appropriate to mathematics are those of intuitionistic logic. Ingeniously, Dummett gives a *general* argument in favor of taking verification conditions as truth conditions – *general* because it does not depend on any special character of mathematical propositions and thus applies everywhere if it applies anywhere. Hence, in this respect, it satisfies our instinctive demand that our theories of meaning and truth (whether or not the latter is based on the former) be uniform across the language, a demand that is made all the more plausible by the fact that the logical vocabulary is the common property of *all* segments of our language: It could be a theoretical embarrassment to be obliged to assign the logical particles different meanings from one subject matter to the next. (What would we do in *mixed* contexts?)

So the generality of his account, and its attendant reconciliation of semantics and epistemology are achieved through a single fundamental move: the founding of the theory of meaning on epistemology – and the theory of truth on this already (for the realist) truncated theory of meaning. The natural result is a theory of truth for propositions, mathematical and nonmathematical alike, in which the *truth-conditions* are the *assertability conditions.* Sentences with no assertability conditions just

do not express propositions – they do not describe possible but possibly unknowable states of the universe.

But one pays a price.

It is the abandonment of classical logic in favor of intuitionistic logic. Of course, to describe it as a *price* is to side with the platonist in the dispute: For the intuitionist there is no cost, simply gains on all fronts, as logic, mathematics, epistemology, semantics, and metaphysics are all finally brought into harmony.

Whatever one thinks of the details of these views, it is important to notice that a number of positions in the philosophy of mathematics (and other branches of philosophy that don't concern us here) are united by precisely such epistemic considerations.

An excellent example of a view that is similar in spirit is to be found in Putnam's "Models and Reality," a piece falling squarely in the pragmatist tradition. Certain authors, in the course of expressing similar concerns, have tended to legislate out of mathematics (and out of the realm of sense altogether as "speculative metaphysics") any proposition that couldn't "in principle" be decided by us.

The lines of demarcation vary from author to author – depending to a large extent on how "in principle" and "decided" are understood. On this spectrum, Putnam does not come off as a "hard-liner." His view amounts to this: Call our "theory of the world" some set of our beliefs, augmented by their logical consequences and indeed corrected and extended by any canons of reasoning, inductive or deductive, that might ever find favor with us. By a famous theorem due to Löwenheim and Skolem, such a theory – far vaster than any theory anyone has ever actually held or conceivably ever could hold – if it is consistent, has among its models (interpretations) ones of every cardinality from $\aleph_0$ on up, as well as others with even more horrifying pathologies. Which, if any, of these models is "the real world"? There is, *for us,* no distinguishing among them; for any distinctions that could be made on the basis of any principles we might hold or observations we might make *have already been taken into account* in constructing the theory (and therefore in selecting the set of models).

Putnam now asks if there is a fact of the matter as to *which* (if any) of the models of this theory (assuming it has some) is "the real world." (Whether the real world is even among the models may be in question for some. Consider the skeptic who feels there are ways in which our best efforts, as outlined above, might have been *mistaken,* thus possibly excluding "the real world" from the model set under consideration.)

His pragmatist answer is no.

Yet, one feels compelled to ask, might we not have been differently constructed – perhaps in such a way that our reconstructed selves made finer-grained distinctions that those we are now able to make or could ever conceivably make? Or simply different distinctions? If there is to be no fact of the matter about *which* of the models of our idealized theory is the real world, then there cannot *be* possible distinctions we cannot actually make – or could not make in Putnam's ideally extended theory.

The naive realist feels short-changed.

Given the way we are, some worlds are indistinguishable by us, and by the idealized possessors of our ideally extended idealized theories. But surely, had we been differently made, different theories might have evolved, and still different theories might have resulted from a yet different process of idealization. If such theories would be possible theories, surely worlds distinguishable in terms of them would be genuinely different, even if not distinguishable by Putnam's theory.

Or so a naive realist might want to reply.

Whatever the merits of such a reply, its bearing on Putnam's conclusions as applied to mathematics is more difficult to assess. One line of argument might be the following: Perhaps a way in which we might have been different involves the ability to "count" uncountable sets, or "compute" uncomputable functions, or the like. Whether, in Russell's famous phrase, such feats are at most "medically impossible" is certainly one of the questions at issue. If the impossibility *is* merely medical – if it resides in the accidents of our genetic makeup – then such feats might not have been impossible. And if they weren't impossible, we would be able to decide issues we are not able to decide. And if these are issues we *might* have been able to decide, they must be real issues, about which there must be a fact of the matter.

Or so the naive realist might urge.

In their curtailment of reality, some deal evenhandedly with all of it (Putnam). Others (intuitionists) address only mathematical reality – perhaps because they see mathematics as our *creation*, or because they see it as subjective in some important way, or... In each case, the result is a limitation of reality to what, in their view, could possibly be known. And in each case, too, it proceeds through an adjustment of the bounds of possible sense: through the theory of meaning.

Of course it is tendentious to describe their doings as a *curtailment*. This is a luxury we allow ourselves only because we have adopted, for expository purposes, the standpoint of the naive realist, the target of many of these views. In any event, whether such "curtailment" is legitimate depends on subtle questions in the philosophy of language – questions not likely to be resolved very soon, at least to the satisfaction of all

combatants, for they are among the most profound that have plagued philosophers ever since the subject began.

To mention but one vexing question that runs through all the views discussed in this section: How may one express the "fact" that certain propositions cannot be expressed – because if they were expressible, we could not decide them and they would then have violated our *epistemic* constraints on sense. It seems that for us to know that they would violate such constraints *on meaning* we must know *on the basis of what they mean* that we could not know them to be true or false if they meant anything at all.

This problem is an ancient one: How can we circumscribe the bounds of sense without stepping outside of them?

7. The iterative conception of set

Consider the following view (which should be thought of as a sort of mathematico-philosophical metaphor, rather than as a serious suggestion): Sets are created by the mind. They are created by the mental act of "collecting" objects (or mental representations of the objects) together. This act can be performed only if the objects collected together are *already* in existence. So, as a result, the members of a set are always *prior* (in time) to the set.

Such a view has a number of attractive immediate consequences. For one thing, there cannot possibly be a collection of *all* sets. (It would have to contain itself, and this would mean that it would have to exist *prior* to itself.) Nor is there any reason why there *must* be a collection of *all* sets that are not members of themselves (this would, in fact, be the collection of all sets, since *no* set is a member of itself on this conception). So the Russell paradox is avoided. On the other hand, if a set has already been created, then, for any condition *P* that refers only to sets already created, there should be a set consisting of all members of the given set satisfying that condition *P*. For what can stop the mind from "collecting" all the members of the given set that satisfy *P*? So many plausible principles of set existence (e.g., existence of a union of any two sets, of an intersection, of an empty set), including the existence of a power set (a set of *all* subsets of a given set), seem to follow from this conception, while no obviously paradox-breeding principle of set existence seems to.

Although the metaphor of sets coming into existence (or being created) *in time* by the human *mind* cannot be taken seriously (even if the mind could create the set of all integers, separately creating each of its subsets would keep the mind rather busy, not to mention collecting these all together into the power set of the set of integers), these advantages of the metaphor can be retained if we suppose that there is *some* relation of "priority" that is transitive, irreflexive, and asymmetrical, and such that the members of any set are always *prior* to the set. Indeed, just such an intuition led Russell to the Vicious Circle Principles and ultimately to the theory of types. Any conception of set in which this figures as a prominent motivating force is today referred to as *the* iterative conception of set (despite the fact that there is more than one such conception). A number of papers in Part IV of this book discuss this conception, explain its great mathematical importance, and explore what the needed relation of "priority" might be.

One such conception is as follows: There is a well-defined totality of *all* sets (giving up the original metaphor of sets "coming into existence" or being constructed by us completely). This totality is not itself a set. Every set belongs to a set of a special kind, called a *rank* by some authors. (These ranks are a modified version of Russell's types, extended into the transfinite.) The ranks are indexed by numbers (including transfinite numbers, or Cantorian "ordinals"); the relation of priority is just the relation "belonging to a smaller rank" (i.e., a rank indexed by a smaller ordinal). The rank indexed by the number zero contains only the empty set (in the case of pure set theory; in a set theory with individuals ("*Urelemente*") we would take the collection of all the individuals as the rank zero). The rank indexed by the successor of any number (or any ordinal) is the power set of the rank indexed by the number. At those transfinite numbers that are not successors (the so-called "limit numbers," of which the least is ω, the ordinal number of the sequence of all integers) we take the union of all the ranks indexed by earlier ordinals to be the rank indexed by the limit ordinal. (For example, rank ω contains all the sets belonging to any *finite* rank. Thus rank ω is the union, in the set-theoretic sense, of ranks $0, 1, 2, \ldots$)

This explanation of the iterative conception makes free use of *numbers* (in fact, of *transfinite* numbers) and of mappings from sets to numbers, notwithstanding the fact that the numbers will eventually be identified with certain sets. This is not objectionable unless we think of the possibility of identifying mathematical objects with sets as of greater epistemological or metaphysical significance than is now customary among either philosophers or mathematicians. There is no reason why we cannot think of numbers, functions, and so on, as objects concerning which we have a certain amount of mathematical theory prior to doing set theory. This prior theorizing enables us to state certain assumptions about sets (e.g., that every set belongs to some rank); and when we see that these assumptions lead to an attractive theory we can adopt it and

work out a formalization of it that seems attractive. Certain of the sets in such a formalization may *end up* being singled out as numbers (ordinals); but this feature has only "elegance" to recommend it, it cannot, on pain of circularity, be claimed to have epistemological significance – at least not by one who views the iterative conception *itself* as having some epistemological force. (Nor need it be viewed as having profound metaphysical significance either, but that is a more complicated matter.)

If we start with a weak (second-order) theory of ordinal numbers, then the axioms of set theory can actually be *derived* from the assumptions that the ranks exist and are related as described. Thus the iterative picture has the pleasing property of *unifying* the axioms of set theory; it gives a "model" such that we can *prove* the axioms to be true in that model (assuming, of course, the existence of the model), and this is clearly preferable to just assuming the various axioms without giving any intuitive picture of how or why they are supposed to be true. Although some logicians (notably Quine) claim that no intuitive justification can be given for the acceptance of set-theoretic axioms after the discovery of the Russell paradox, the fact is that almost all of our authors (Gödel, Wang, Boolos, and Parsons, among others) maintain that this iterative picture *is* an intuitive justification for the standard axioms in the sense that (1) the picture is natural and persuasive (i.e., it seems to these authors that there is *a* notion of set that we had all along on which the elements of a set had to be "prior" to the set, and on which a "set of all sets" was impossible); and (2) as just remarked, the assumption of a structure of "ranks" with the properties we mentioned enables us to derive almost all of the standard axioms of set theory. (The one doubtful case being the Axiom of Replacement – the image of a set under a function is a set – although that, too, seems natural to many people under this picture even if it is not deducible from the hypothesis that the system of ranks exists.)

The version of the iterative conception just described is "platonistic" in that it views the sets as all existing at once. The priority relation is simply an ordering defined in terms of the membership relation of set theory itself. Others espouse the iterative conception but seek to avoid assuming a well-defined totality of all sets. Saul Kripke has suggested that this might be made sense of by construing set-theoretical quantifiers intuitionistically rather than classically. On such a view it is not clear why a condition P that quantifies over *all* sets ("all" read intuitionistically) should single out a well-defined subset of a *given* set, however; so this construal might leave us with a problem in justifying or even stating the Axiom of Selection. In an interesting article in Part IV, Charles Parsons proposes a *modal* interpretation of the "priority" relation on which the iterative conception rests. (The idea being that the members of a set *could exist* – in some possible well-founded structure that is thought of as a "realization" of set theory in a possible world – without the set existing. The priority relation is a kind of *presupposition* relation; x is prior to y if y's existence *presupposes* x's existence, but not vice versa.)

There is no doubt that the iterative conception connects with a powerful and useful mathematical metaphor. On the other hand, the large number of statements to the effect that *the* iterative conception is a perfectly clear and consistent conception that shows there is no difficulty at all with our set-theoretic "intuition" might suggest to some readers that "the lady doth protest too much." The problem in either justifying or doing without the assumption that all sets form a well-defined totality on the iterative conception, and the *epistemological* unclarity of the "priority" relation, suggest that the iterative conception is not without its own problems.

8. The problem of "access"

At one extreme of the spectrum of views one might hold on the foundations of set theory, and of mathematics generally, stands a form of "platonism" that might be described as follows. Mathematics consists of a body of propositions about an independent reality composed of the familiar mathematical objects (such as sets, numbers, functions, and spaces). Mathematical *discovery* is the uncovering of truths about this independently existing reality by deduction from axioms that we see to be true by a special faculty of intuition distinct from sense experience (which gives us knowledge only of the empirical world). Mathematical objects are independent of our minds and, unlike physical objects, do not interact with our bodies to cause alterations in our brains that lead ultimately to knowledge of them. But they must be postulated to account for the existence and growth of mathematical knowledge and, to the extent to which other knowledge is dependent on mathematical knowledge, of other knowledge as well.

It is hard to pin this view in its pure form on anyone, although Gödel perhaps comes as close to it as anyone since Plato. For the time being, we will consider the view in the abstract, returning later to the details of the view that Gödel expresses in his various writings.

It is likely that most mathematicians would reject this extreme form of platonism; certainly few contemporary philosophers or psychologists would find the idea of a nonphysical power of surveying a realm of independently existing objects outside of space and time very congenial. But if one rejects the idea of returning to such a view (a view that is "platonist" in a more literal sense than that overused epithet frequently has in philosophical discussion), then one cannot escape the problems posed by the talk of "intuition" that one encounters in the writings of many philosophers and set theorists. What is "intuition," and how can there be such a faculty if platonist views such as those mentioned above are totally wrong?

We might describe the problem as a problem of *access: here* are we, evolving social organisms in space-time. Our sense organs are admirably suited to bringing us information about tables and chairs, trees, fruits and vegetables, other organisms, the sky, the weather, and so on. We have managed to devise electronic and optical extensions of these sense organs that enable us to observe objects as small as viruses (and even smaller) and as distant as remote galaxies. But none of these sense organs, natural or artificial, extended or unextended, ever causally interacts with, observes, or perceives a *set*. *There* are the sets; beautiful (at least to some), imperishable, multitudinous, intricately connected. They toil not, neither do they spin. Nor, and this is the rub, do they interact with us in any way. So how are we supposed to have *epistemological access* to them? To answer, "by intuition," is hardly satisfactory. We need some account of how we can have *knowledge* of these beasties, some account of our cognitive relationship to them.

(We referred earlier to worries about whether the notion of a set is "clear." It seems to us that many such worries are really grounded in precisely this problem of access. It is not, after all, that set language is "unclear" in any ordinary linguistic sense: too many ambiguities, too many possible paraphrases, and the like. What is "unclear" is whether such objects really exist and, if they do, how we can possibly know what we claim to know about them.)

One suggestion that a number of authors have advanced is that sets may exist not as platonic, extra-mental objects, but *in the mind*, as objects of our own making. If sets are, in some way, our own creation and in our minds, then the problem of accounting for our access to them should be easier. Or so it might seem.

But how can a nondenumerable infinity of sets exist as mental objects in our all-too-finite minds? One possible answer would be that they exist as *intentional objects*, that is, as objects whose existence is the *content* of certain thoughts of ours. These thoughts need not be supposed to be true of anything "external" any more than the play *Hamlet* is true of any actual prince of Denmark; truth in set theory, on such an account, would be no more than *truth in the story*. Two problems with such an account (which, along with Gödel's account and others, is discussed by Wang in Part IV) are:

(1) On such an account, how are we to choose among rival set theories? Or should we just be relativists? Should we say that within broad limits, one set theory is as good as another? And if so, what would establish even the broad limits? Wang points out that such an attitude is anathema to most working set theorists. He suggests (tentatively) that perhaps truth in set theory might be defined in terms of convergence in the intuitions of set theorists in the long run. But is such convergence, assuming it exists at all, epistemologically significant if it is founded on "aesthetic" preferences for certain stories as opposed to others? And how much of the existing agreement among set theorists is just the product of academic fashion?

(2) How, on such an account, can we explain the apparent truth or approximate truth of empirical laws (e.g., Newton's Law of Gravity, Quantum Mechanics, Relativity) that require higher mathematics to state? Such a law, as standardly formalized, makes reference both to physical entities (forces, masses, particles, fields) and to functions and sets. If the functions and sets are just *intentional* objects, objects "in the story," then the physical theories are to that extent also about fictional (or at least intentional) objects, too. This view would seem to require a radical adjustment in philosophy of science as a whole, not just a new philosophy of mathematics. If mathematics is fiction, should the Mathematics Department be renamed "Creative Writing"? And what about the Physics Department? (Perhaps all there is is Creative Writing.)

Parsons's paper, which is related to Putnam's "Mathematics without Foundations," explores the possibility of a modal interpretation of set theory. (Parsons is continuing to explore this possibility in subsequent work that is not yet published as of this writing.) The idea is that set theory should be interpreted as a theory of what sorts of structures *could* exist (in a special mathematical sense of "could") and, in particular, as a theory of what models for iterative set theory *could* exist.

One question of mathematical interest raised by this program is the following: Is there some way of making the Axiom of Replacement (the range of values of any function whose domain is a set is also a set) more evident by deriving it from suitable assumptions about possible existence? Are there axioms that are "evident" on the modal interpretation that imply the Axiom of Replacement?

At first blush, the modal logical interpretation seems attractive because it eschews the whole picture that makes the problem of "access" seem so terrifying. It avoids saying that there is a nondenumerable infinity of actual objects called sets *there*, thus bypassing the problem of how we *here* can know about those objects *there* if there are no telegraph (or other) wires running from *there* to *here*. (It also rejects Quine's view that

we should reformulate all scientific theories in quantificational logic in order to determine their "ontological commitments." Recasting set theory in *modal* logic leaves it unclear what the "ontological commitments" of set theory are, since there is no presently agreed upon account of ontological commitment in the case of a logic with modal operators. But if one thinks that the whole picture is wrong, then one will not regard this as a defect of the reformulation.) The problem of accounting for our epistemic access to a nondenumerable infinity of *recherché* entities is replaced with the problem of accounting for our ability to know modal truths, truths about what is and is not possible. And perhaps this will prove in the long run to be more tractable; to some it sounds less frightening.

Be that as it may, at present it is not clear just *how* the modal logical interpretation *can* help with the epistemological problem. This is particularly so if modalities are themselves understood in terms of "possible world" semantics: S is mathematically necessary if it is true in every mathematically possible world; S is mathematically possible if S is true in at least one mathematically possible world. The problem then shifts to that of explaining how we can know what we appear to know about this new breed of platonic entity – the (mathematically) possible world. After all, the theories that naturally come to mind as candidate accounts of our ability to know modal truths are just the ones that come to mind on the standard "Mathematics as the theory of Mathematical Objects" picture: for example, that modal truths are "analytic" or "true by convention"; that there are special Faculties of the Mind that enable one to know (and perhaps actually *define*) what is and is not "possible" in the logical/mathematical sense; that our theory of what is and is not *mathematically* possible is, like our theory of what is and is not *physically* possible, a part of total scientific theory and to be accepted, modified, or rejected on grounds similar to those on which one accepts, modifies, or rejects empirical theories. It may be useful to have an alternative to the Heaven of Mathematical Objects picture; but we haven't been shown (yet) *how* it is useful.

9. Quine and Gödel

In line with what we said at the outset about the importance of seeing connections between issues raised by authors who are philosophers and authors who are mathematicians, and the importance of seeing connections between essays in different sections of this anthology, we recommend reading Quine's "Carnap and Logical Truth" in connection with the papers in Part IV as well as in connection with the other papers in

Part III. In particular, the view just mentioned, that the grounds for accepting or rejecting mathematical theories are analogous to the grounds for accepting or rejecting physical theories, has long been urged by Quine.

It is not, of course, that Quine is unaware of the fact that there are *experiments* in physics and no experiments in mathematics (or, at least, none in the same sense). But Quine emphasizes the idea that mathematics has to be viewed not by itself but rather as a part of an all-embracing conceptual scheme, and he claims that the necessity for quantification over mathematical objects (e.g., *sets*) is all the reason one needs for making the "posit" of the existence of sets, numbers, and so on. Sets and electrons are alike for Quine in being objects we need to assume to do science.

While this sort of holistic pragmatism is attractive in that it recognizes what Russell and the logicists were for a long time alone in emphasizing – that we must account for the use of mathematical locutions in empirical statements, and not only in the context of pure set theory, number theory, and so on – and in that it provides a reason to believe in the existence of sets without postulating mysterious Faculties of the Mind, it too runs into serious difficulties. (So does *every* view; that is why the philosophy of mathematics is so fascinating.) Quine *seems* to be saying (on one reading; despite the deceptive clarity of his style, he is not always an easy philosopher to interpret) that science as a whole is to be viewed as a single explanatory theory and that the theory is to be justified *as a whole* by its ability to explain *sensations*. But is is not clear what the acceptance or nonacceptance of the Axiom of Choice or the Continuum Hypothesis has to do with explaining sensations.

The idea that there is something *analogous* to empirical reasoning in pure mathematics has also been advanced by Putnam ("Mathematical Truth," not in this volume) and even by Gödel ("What is Cantor's Continuum Problem?" in Part IV). In spite of his platonism, Gödel is much too sophisticated to think acts of "intuition" are *all* that is involved in mathematical "self-evidence," "plausibility," and the like. The fact that two philosophers as radically different as Quine and Gödel both recognize the presence of an element of something like "hypothetico-deductive" reasoning in pure mathematics is certainly striking. (Such an element was also pointed out by Russell, in the preface to *PM* and in earlier publications.)

Quine recognizes that even in empirical science there are considerations other than predicting sensations that play a crucial role in theory selection. He speaks most often of "conservatism" – the desire to preserve principles that have long been regarded as "central" or "obvious" or both, and of "simplicity" – a desire for elegance, which occasionally

makes us fly in the face of conservatism when a radical change of a central (or even a "self-evident") principle turns out to lead to far-reaching simplifications of the whole system.

But why should the simplest and most conservative system (or rather, the system that best balances simplicity and conservatism, by our lights) have any tendency to be *true*? Quine, good pragmatist that he is, tends to pooh-pooh this sort of question; but more realistically minded philosophers are sure to be bothered. It is hard enough to believe that the natural world is so nicely arranged that what is simplest, etc., by *our* lights is always the same as what is *true* (or, at least, *generally* the same as what is true); why should one believe that the universe of sets (or the totality of modal truths) is so nicely arranged that there is a preestablished harmony between *our* feelings of simplicity, etc., and *truth*?

It might be rewarding at this point to go into Gödel's view in more detail and particularly to compare it to Quine's; for despite superficial differences, they agree in a surprising number of respects. And what is perhaps of greater interest, the least satisfactory portion of each account comes at precisely the same spot: at the place where it must be explained how the criteria of truth that are advanced by each are related to the truth of the propositions for whose truths they are criteria.

For Gödel, as for Quine, objects exist in exactly the same sense as the objects of physical theory. And our reasons for believing in them are every bit as good as our reasons for believing in the existence of fields, protons, and so on. Both Gödel and Quine declare our experience to be in some sense the touchstone of our theorizing; they differ principally in what they admit as constituting that experience. Quine insists on an almost thoroughgoing holism, broken only by the independence of observation sentences (they record our experience) and (sometimes) logical truths, whereas Gödel adopts a modified Kantian position about experience: Our experience of physical objects goes beyond mere sensation – our concept of physical object contains an admixture of elements not conceivably derived from sensation. But the "added ingredients" belong to our perceptual faculty and, unlike for Kant, they are not subjective, not contributed by the perceiving subject ("... by our thinking we cannot create any qualitatively new elements, but only reproduce and combine those that are given"; p. 484). This faculty of grouping sensations into sensations *of objects* is, for Gödel, very closely related to the faculty of intuition, in virtue of which we have the (iterative) concept of set ("... the function of both is 'synthesis,' i.e., the generating of unities out of manifolds...."; p. 484, fn. 26). He emphasizes that it (intuition) need not be conceived of as giving us *immediate* knowledge of mathematics. But this should not exclude it from being recognized as part of experience, particularly since even our experience of physical objects transcends sensation. He concludes: "Evidently the 'given' underlying mathematics is closely related to the abstract elements contained in our empirical ideas" (p. 484).

Despite this difference (and who can say how much of a difference it really is, since Quine much more readily discusses the kinds of sentences we may admit into our theories than the kinds of evidence appropriate to each?), both agree that "empirical" considerations might figure among the criteria for the truth of mathematical axioms. According to Quine's view, this is axiomatic, since mathematical axioms are not observation sentences, while Gödel concedes that "besides mathematical intuition, there exists another (though only probable) criterion of the truth of mathematical axioms, namely their fruitfulness in mathematics and, one may add, possibly also in physics" (p. 485). Someone wishing to minimize the difference would emphasize that Gödel's account of the connection between mathematical experience and our acceptance of a mathematical theory plays a similar *explanatory* role in his philosophy to that played by the concept of simplicity, conservatism, and explanatory power in the philosophies of pragmatists and empiricists: In neither case is it made crystal clear why theories that (1) accord with intuition or (2) appear simpler and appear to explain better are more likely to be true than the others. Far from being insensitive to the problem of "access," as we have described it in this section, Gödel rather founds his philosophy on the view that it must be taken very seriously indeed. Believing that we *have* peculiarly mathematical knowledge, he tries to explain our possession of it by (1) noting that there are objective elements of experience that do not derive from sensation and (2) proposing to account for our mathematical knowledge at least in part in terms of such nonsensational aspects of experience. To be sure, we are left without an account of the *mechanism* by which these experiential elements reflect aspects of the alleged reality they allegedly betray. But it would be hard to argue that such a view is in a worse position for giving an account of our overall knowledge than are its holist, empiricist, or pragmatist competitors. In both cases, the philosophical lacuna occurs in exactly the same place: when it must be explained why the criteria that are advocated for selecting a theory (fit with intuition – simplicity) should pick out a theory that is more likely to be true than one of its competitors.

An interesting and perhaps undesired consequence of Gödel's view, and one that makes the *rapprochement* with Quine even closer, is that mathematics no longer appears to be a priori, unless that is reinterpreted as "independent of *sensation*." For now, our only mathematical knowledge is what we have derived from our experience. And Gödel dispenses

with the very feature of the Kantian view that was designed to guarantee the a priori character of mathematics when he says:

> It by no means follows, however, that the data of this second kind [what we have called the nonsensational component of experience. Eds.], because they cannot be associated with actions of certain things upon our sense organs, are something purely subjective, as Kant asserted. Rather they, too, may represent an aspect of objective reality, but, as opposed to the sensations, their presence may be due to another kind of relationship between ourselves and reality. (p. 484)

The very question "Why should one believe that there is a preestablished harmony between our feelings of simplicity, or intuition, and *truth*?" presupposes a notion of truth that is independent of our standards of assertability. In Dummett's terminology, it assumes a "realist" notion of truth; and it is at this point that the discussion of issues in the foundations of set theory leads back to the discussion of the issues that we grouped together in Section 6 under the heading "Mathematical Truth." It is possible that significant further progress on these issues cannot be made until we have a more satisfactory account of the nature of truth and of the ways in which truth and reference are and are not linked to assertability. In any case, there is certainly a close connection between discussions in the philosophy of mathematics and discussions in general philosophy concerning the metaphysical issue of realism. But this is not to say that philosophy of mathematics ought simply to wait for improved views in general philosophy. Quite the contrary, for it is even more likely that one way in which theories of truth and knowledge in general philosophy will be shown to be adequate (or inadequate) is by their ability (or inability) to account for mathematical knowledge; and it is only in philosophy of mathematics that one finds searching attempts to apply theories of truth and knowledge to the special case of mathematics. For this reason, it is our conviction that philosophers interested in fundamental questions in epistemology, in theory of reference and truth, in philosophy of language, and in metaphysics should pay closer attention to the case of mathematics than they generally have. General philosophical theories that appear to account adequately for our intercourse with protons and pachyderms but fail on polynomials are, for that very reason, inadequate treatments of the very cases they seem to fit so well. The world of mathematics is not a world apart. We will not have an adequate account of the physical world and our knowledge of it until we understand better than we presently do the role played by mathematics in our accounts of physical phenomena. And it is not likely that we will have satisfied ourselves on that score until we have produced accounts of knowledge, truth, and reality that deal adequately with pure mathematics as well.

PART I

The foundations of mathematics

Symposium on the foundations of mathematics

1. The logicist foundations of mathematics

RUDOLF CARNAP

The problem of the logical and epistemological foundations of mathematics has not yet been completely solved. This problem vitally concerns both mathematicians and philosophers, for any uncertainty in the foundations of the "most certain of all the sciences" is extremely disconcerting. Of the various attempts already made to solve the problem none can be said to have resolved every difficulty. These efforts, the leading ideas of which will be presented in these three papers, have taken essentially three directions: *Logicism*, the chief proponent of which is Russell; *Intuitionism*, advocated by Brouwer; and Hilbert's *Formalism*.

Since I wish to draw you a rough sketch of the salient features of the logicist construction of mathematics, I think I should not only point out those areas in which the logicist program has been completely or at least partly successful but also call attention to the difficulties peculiar to this approach. One of the most important questions for the foundations of mathematics is that of the relation between mathematics and logic. *Logicism* is the thesis that mathematics is reducible to logic, hence nothing but a part of logic. Frege was the first to espouse this view (1884). In their great work, *Principia Mathematica*, the English mathematicians A. N. Whitehead and B. Russell produced a systematization of logic from which they constructed mathematics.

We will split the logicist thesis into two parts for separate discussion:

1. The *concepts* of mathematics can be derived from logical concepts through explicit definitions.
2. The *theorems* of mathematics can be derived from logical axioms through purely logical deduction.

I. The derivation of mathematical concepts

To make precise the thesis that the concepts of mathematics are derivable from logical concepts, we must specify the logical concepts to be employed

The first three essays in this chapter form part of a symposium on the foundations of mathematics which appeared in Erkenntnis (1931), pp. 91–121. They were translated by Erna Putnam and Gerald J. Massey and appear here with the kind permission of Rudolf Carnap, Arend Heyting, and Klara von-Neumann Eckart. The last of these appears in A. H. Taub, ed., *John von Neumann Collected Works*, Vol. 2 (New York: Pergamon Press, 1961).

in the derivation. They are the following: In propositional calculus, which deals with the relations between unanalyzed sentences, the most important concepts are: the negation of a sentence p, 'not-p' (symbolized '$\sim p$'); the disjunction of two sentences, 'p or q' ('$p \vee q$'); the conjunction, 'p and q' ('$p \cdot q$'); and the implication, 'if p, then q' ('$p \supset q$'). The concepts of functional calculus are given in the form of functions, e.g., '$f(a)$' (read 'f of a') signifies that the property f belongs to the object a. The most important concepts of functional calculus are universality and existence: '$(x)f(x)$' (read 'for every x, f of x') means that the property f belongs to every object; '$(\exists x)f(x)$' (read 'there is an x such that f of x') means that f belongs to at least one object. Finally there is the concept of identity: '$a=b$' means that 'a' and 'b' are names of the same object.

Not all these concepts need be taken as undefined or primitive, for some of them are reducible to others. For example, '$p \vee q$' can be defined as '$\sim(\sim p \cdot \sim q)$' and '$(\exists x)f(x)$' as '$\sim(x) \sim f(x)$'. It is the logicist thesis, then, that the logical concepts just given suffice to define all mathematical concepts, that over and above them no specifically mathematical concepts are required for the construction of mathematics.

Already before Frege, mathematicians in their investigations of the interdependence of mathematical concepts had shown, though often without being able to provide precise definitions, that all the concepts of arithmetic are reducible to the natural numbers (i.e., the numbers 1, 2, 3, ... which are used in ordinary counting). Accordingly, the *main problem* which remained for logicism was to derive the natural numbers from logical concepts. Although Frege had already found a solution to this problem, Russell and Whitehead reached the same results independently of him and were subsequently the first to recognize the agreement of their work with Frege's. The crux of this solution is the correct recognition of the logical status of the natural numbers; they are logical attributes which belong, not to things, but to concepts. That a certain number, say 3, is the number of a concept means that three objects fall under it. We can express the very same thing with the help of the logical concepts previously given. For example, let '$2_m(f)$' mean that at least two objects fall under the concept f. Then we can define this concept as follows (where '$=_{\mathrm{Df}}$' is the symbol for definition, read as "means by definition"):

$$2_m(f) =_{\mathrm{Df}} (\exists x)(\exists y)[\sim(x=y) \cdot f(x) \cdot f(y)]$$

or in words: there is an x and there is a y such that x is not identical with y and f belongs to x and f belongs to y. In like manner, we define 3_m, 4_m, and so on. Then we define the number two itself thus:

$$2(f) =_{\mathrm{Df}} 2_m(f) \cdot \sim 3_m(f)$$

or in words: at least two, but not at least three, objects fall under f. We can also define arithmetical operations quite easily. For example, we can define addition with the help of the disjunction of two mutually exclusive concepts. Furthermore, we can define the concept of natural number itself.

The derivation of the other kinds of number – i.e., the positive and negative numbers, the fractions, the real and the complex numbers – is accomplished, not in the usual way by adding to the domain of the natural numbers, but by the construction of a completely new domain. The natural numbers do not constitute a subset of the fractions but are merely correlated in obvious fashion with certain fractions. Thus the natural number 3 and the fraction 3/1 are not identical but merely correlated with one another. Similarly we must distinguish the fraction 1/2 from the real number correlated with it. In this paper, we will treat only the definition of the real numbers. Unlike the derivations of the other kinds of numbers which encounter no great difficulties, the derivation of the real numbers presents problems which, it must be admitted, neither logicism, intuitionism, nor formalism has altogether overcome.

Let us assume that we have already constructed the series of fractions (ordered according to magnitude). Our task, then, is to supply definitions of the real numbers based on this series. Some of the real numbers, the rationals, correspond in obvious fashion to fractions; the rest, the irrationals, correspond as Dedekind showed (1872) to "gaps" in the series of fractions. Suppose, for example, that we divide the (positive) fractions into two classes, the class of all whose square is less than 2, and the class comprising all the rest of the fractions. This division forms a "cut" in the series of fractions which corresponds to the irrational real number $\sqrt{2}$. This cut is called a "gap" since there is no fraction correlated with it. As there is no fraction whose square is two, the first or "lower" class contains no greatest member, and the second or "upper" class contains no least member. Hence, to every real number there corresponds a cut in the series of fractions, each irrational real number being correlated with a gap.

Russell developed further Dedekind's line of thought. Since a cut is uniquely determined by its "lower" class, Russell defined a real number as the lower class of the corresponding cut in the series of fractions. For example, $\sqrt{2}$ is defined as the class (or property) of those fractions whose square is less than two, and the rational real number 1/3 is defined as the class of all fractions smaller than the fraction 1/3. On the basis of these definitions, the entire arithmetic of the real numbers can be developed. This development, however, runs up against certain difficulties connected with so-called "impredicative definition," which we will discuss shortly.

The essential point of this method of introducing the real numbers is that they are *not postulated but constructed.* The logicist does not establish the existence of structures which have the properties of the real numbers by laying down axioms or postulates; rather, through explicit definitions, he produces logical constructions that have, by virtue of these definitions, the usual properties of the real numbers. As there are no "creative definitions," definition is not creation but only name-giving to something whose existence has already been established.

In similarly constructivistic fashion, the logicist introduces the rest of the concepts of mathematics, those of analysis (e.g., convergence, limit, continuity, differential, quotient, integral, etc.) and also those of set theory (notably the concepts of the transfinite cardinal and ordinal numbers). This "constructivist" method forms part of the very texture of logicism.

II. The derivation of the theorems of mathematics

The second thesis of logicism is that the *theorems of mathematics* are derivable from logical axioms through logical deduction. The requisite system of logical axioms, obtained by simplifying Russell's system, contains four axioms of propositional calculus and two of functional calculus. The rules of inference are a rule of substitution and a rule of implication (the *modus ponens* of ancient logic). Hilbert and Ackermann have used these same axioms and rules of inference in their system.

Mathematical predicates are introduced by explicit definitions. Since an explicit definition is nothing but a convention to employ a new, usually much shorter, way of writing something, the *definiens* or the new way of writing it can always be eliminated. Therefore, as every sentence of mathematics can be translated into a sentence which contains only the primitive logical predicates already mentioned, this second thesis can be restated thus: Every provable mathematical sentence is translatable into a sentence which contains only primitive logical symbols and which is provable in logic.

But the derivation of the theorems of mathematics poses certain difficulties for logicism. In the first place it turns out that some theorems of arithmetic and set theory, if interpreted in the usual way, require for their proof besides the logical axioms still other special axioms known as the *axiom of infinity* and the *axiom of choice* (or multiplicative axiom). The axiom of infinity states that for every natural number there is a greater one. The axiom of choice states that for every set of disjoint nonempty sets, there is (at least) one selection-set, i.e., a set that has exactly one member in common with each of the member sets. But we are not concerned here with the content of these axioms but with their logical

character. Both are existential sentences. Hence, Russell was right in hesitating to present them as logical axioms, for logic deals only with possible entities and cannot make assertions about whether something does or does not exist. Russell found a way out of this difficulty. He reasoned that since mathematics was also a purely formal science, it too could make only conditional, not categorical, statements about existence: if certain structures exist, then there also exist certain other structures whose existence follows logically from the existence of the former. For this reason he transformed a mathematical sentence, say S, the proof of which required the axiom of infinity, I, or the axiom of choice, C, into a conditional sentence; hence S is taken to assert not S, but $I \supset S$ or $C \supset S$, respectively. This conditional sentence is then derivable from the axioms of logic.

A greater difficulty, perhaps the greatest difficulty, in the construction of mathematics has to do with another axiom posited by Russell, the so-called *axiom of reducibility,* which has justly become the main bone of contention for the critics of the system of *Principia Mathematica.* We agree with the opponents of logicism that it is inadmissible to take it as an axiom. As we will discuss more fully later, the gap created by the removal of this axiom has certainly not yet been filled in an entirely satisfactory way. This difficulty is bound up with Russell's *theory of types* which we shall now briefly discuss.

We must distinguish between a "simple theory of types" and a "ramified theory of types." The latter was developed by Russell but later recognized by Ramsey to be an unnecessary complication of the former. If, for the sake of simplicity, we restrict our attention to one-place functions (properties) and abstract from many-place functions (relations), then type theory consists in the following classification of expressions into different "types": To type 0 belong the names of the objects ("individuals") of the domain of discourse (e.g., $a, b, \ldots$). To type 1 belong the properties of these objects (e.g., $f(a), g(a), \ldots$). To type 2 belong the properties of these properties (e.g., $F(f), G(f), \ldots$); for example, the concept $2(f)$ defined above belongs to this type. To type 3 belong the properties of properties of properties, and so on. The basic rule of type theory is that every predicate belongs to a determinate type and can be meaningfully applied only to expressions of the next lower type. Accordingly, sentences of the form $f(a), F(f), 2(f)$ are always meaningful, i.e., either true or false; on the other hand combinations like $f(g)$ and $f(F)$ are neither true nor false but meaningless. In particular, expressions like $f(f)$ or $\sim f(f)$ are meaningless, i.e., we cannot meaningfully say of a property either that it belongs to itself or that it does not. As we shall see, this last result is important for the elimination of the antinomies.

This completes our outline of the simple theory of types, which most proponents of modern logic consider legitimate and necessary. In his system, Russell introduced the ramified theory of types, which has not found much acceptance. In this theory the properties of each type are further subdivided into "orders." This division is based, not on the kind of objects to which the property belongs, but on the form of the definition which introduces it. Later we shall consider the reasons why Russell believed this further ramification necessary. Because of the introduction of the ramified theory of types, certain difficulties arose in the construction of mathematics, especially in the theory of real numbers. Many fundamental theorems not only could not be proved but could not even be expressed. To overcome this difficulty, Russell had to use brute force; i.e., he introduced the axiom of reducibility by means of which the different orders of a type could be reduced in certain respects to the lowest order of the type. The sole justification for this axiom was the fact that there seemed to be no other way out of this particular difficulty engendered by the ramified theory of types. Later Russell himself, influenced by Wittgenstein's sharp criticism, abandoned the axiom of reducibility in the second edition of *Principia Mathematica* (1925). But, as he still believed that one could not get along without the ramified theory of types, he despaired of the situation. Thus we see how important it would be, not only for logicism but for any attempt to solve the problems of the foundations of mathematics, to show that the simple theory of types is sufficient for the construction of mathematics out of logic. A young English mathematician and pupil of Russell, Ramsey (who unfortunately died this year, i.e., 1930), in 1926 made some efforts in this direction which we will discuss later.

III. The problem of impredicative definition

To ascertain whether the simple theory of types is sufficient or must be further ramified, we must first of all examine the reasons which induced Russell to adopt this ramification in spite of its most undesirable consequences. There were two closely connected reasons: the necessity of eliminating the logical antinomies and the so-called "vicious circle" principle. We call "logical antinomies" the contradictions which first appeared in set theory (as so-called "paradoxes") but which Russell showed to be common to all logic. It can be shown that these contradictions arise in logic if the theory of types is not presupposed. The simplest antinomy is that of the concept "impredicable." By definition a property is "impredicable" if it does not belong to itself. Now is the property "impredicable" itself impredicable? If we assume that it is, then since it belongs to

itself it would be, according to the definition of "impredicable," not impredicable. If we assume that it is not impredicable, then it does not belong to itself and hence, according to the definition of "impredicable," is impredicable. According to the law of excluded middle, it is either impredicable or not, but both alternatives lead to a contradiction. Another example is Grelling's antinomy of the concept "heterological." Except that it concerns predicates rather than properties, this antinomy is completely analogous to the one just described. By definition, a predicate is "heterological" if the property designated by the predicate does not belong to the predicate itself. (For example, the word 'monosyllabic' is heterological, for the word itself is not monosyllabic.) Obviously both the assumption that the word 'heterological' is itself heterological as well as the opposite assumption lead to a contradiction. Russell and other logicians have constructed numerous antinomies of this kind.

Ramsey has shown that there are two completely different kinds of antinomies. Those belonging to the first kind can be expressed in logical symbols and are called "logical antinomies" (in the narrower sense). The "impredicable" antinomy is of this kind. Ramsey has shown that this kind of antinomy is eliminated by the simple theory of types. The concept "impredicable," for example, cannot even be defined if the simple theory of types is presupposed, for an expression of the form, a property does not belong to itself ($\sim f(f)$), is not well-formed, and meaningless according to that theory.

Antinomies of the second kind are known as "semantical" or "epistemological" antinomies. They include our previous example, "heterological," as well as the antinomy, well-known to mathematicians, of the smallest natural number which cannot be defined in German with fewer than 100 letters. Ramsey has shown that antinomies of this second kind cannot be constructed in the symbolic language of logic and therefore need not be taken into account in the construction of mathematics from logic. The fact that they appear in word languages led Russell to impose certain restrictions on logic in order to eliminate them, viz., the ramified theory of types. But perhaps their appearance is due to some defect of our ordinary word language.

Since antinomies of the first kind are already eliminated by the simple theory of types and those of the second kind do not appear in logic, Ramsey declared that the ramified theory of types and hence also the axiom of reducibility were superfluous.

Now what about Russell's second reason for ramifying the theory of types, viz., the vicious circle principle? This principle, that "no whole may contain parts which are definable only in terms of that whole", may also be called an "injunction against impredicative definition." A definition is said to be "impredicative" if it defines a concept in terms of a totality to which the concept belongs. (The concept "impredicative" has nothing to do with the aforementioned pseudo concept "impredicable.") Russell's main reason for laying down this injunction was his belief that antinomies arise when it is violated. From a somewhat different standpoint Poincaré before, and Weyl after, Russell also rejected impredicative definition. They pointed out that an impredicatively defined concept was meaningless because of the circularity in its definition. An example will perhaps make the matter clearer:

We can define the concept "inductive number" (which corresponds to the concept of natural number including zero) as follows: A number is said to be "inductive" if it possesses all the hereditary properties of zero. A property is said to be "hereditary" if it always belongs to the number $n+1$ whenever it belongs to the number n. In symbols,

$$\mathrm{Ind}(x) =_{\mathrm{Df}} (f)[(\mathrm{Her}(f) \cdot f(0)) \supset f(x)]$$

To show that this definition is circular and useless, one usually argues as follows: In the *definiens* the expression '(f)' occurs, i.e., "for all properties (of numbers)". But since the property "inductive" belongs to the class of all properties, the very property to be defined already occurs in a hidden way in the *definiens* and thus is to be defined in terms of itself, an obviously inadmissible procedure. It is sometimes claimed that the meaninglessness of an impredicatively defined concept is seen most clearly if one tries to establish whether the concept holds in an individual case. For example, to ascertain whether the number three is inductive, we must, according to the definition, investigate whether every property which is hereditary and belongs to zero also belongs to three. But if we must do this for every property, we must also do it for the property "inductive" which is also a property of numbers. Therefore, in order to determine whether the number three is inductive, we must determine among other things whether the property "inductive" is hereditary, whether it belongs to zero, and finally – this is the crucial point – whether it belongs to three. But this means that it would be impossible to determine whether three is an inductive number.

Before we consider how Ramsey tried to refute this line of thought, we must get clear about how these considerations led Russell to the ramified theory of types. Russell reasoned in this way: Since it is inadmissible to define a property in terms of an expression which refers to "all properties," we must subdivide the properties (of type 1): To the "first order" belong those properties in whose definition the expression 'all properties' does not occur; to the "second order" those in whose definition the expression 'all properties of the first order' occurs; to the "third order"

those in whose definition the expression 'all properties of the second order' occurs, and so on. Since the expression 'all properties' without reference to a determinate order is held to be inadmissible, there never occurs in the definition of a property a totality to which it itself belongs. The property "inductive," for example, is defined in this no longer impredicative way: A number is said to be "inductive" if it possess all the hereditary properties of the first order which belong to zero.

But the ramified theory of types gives rise to formidable difficulties in the treatment of the real numbers. As we have already seen, a real number is defined as a class, or what comes to the same thing, as a property of fractions. For example, we say that $\sqrt{2}$ is defined as the class or property of those fractions whose square is less than two. But since the expression 'for all properties' without reference to a determinate order is inadmissible under the ramified theory of types, the expression 'for all real numbers' cannot refer to all real numbers without qualification but only to the real numbers of a determinate order. To the first order belong those real numbers in whose definition an expression of the form 'for all real numbers' does not occur; to the second order belong those in whose definition such an expression occurs, but this expression must be restricted to "all real numbers of the first order," and so on. Thus there can be neither an admissible definition nor an admissible sentence which refers to all real numbers without qualification.

But as a consequence of this ramification, many of the most important definitions and theorems of real number theory are lost. Once Russell had recognized that his earlier attempt to overcome it, viz., the introduction of the axiom of reducibility, was itself inadmissible, he saw no way out of this difficulty. The *most difficult problem* confronting contemporary studies in the foundations of mathematics is this: How can we develop logic if, on the one hand, we are to avoid the danger of the meaninglessness of impredicative definitions and, on the other hand, are to reconstruct satisfactorily the theory of real numbers?

IV. Attempt at a solution

Ramsey (1926a) outlined a construction of mathematics in which he courageously tried to resolve this difficulty by declaring the forbidden impredicative definitions to be perfectly admissible. They contain, he contended, a circle but the circle is harmless, not vicious. Consider, he said, the description 'the tallest man in this room'. Here we describe something in terms of a totality to which it itself belongs. Still no one thinks this description inadmissible since the person described already exists and is only singled out, not created, by the description. Ramsey believed that the same considerations applied to properties. The totality of properties already exists in itself. That we men are finite beings who cannot name individually each of infinitely many properties but can describe some of them only with reference to the totality of all properties is an empirical fact that has nothing to do with logic. For these reasons Ramsey allows impredicative definition. Consequently, he can both get along with the simple theory of types and still retain all the requisite mathematical definitions, particularly those needed for the theory of the real numbers.

Although this happy result is certainly tempting, I think we should not let ourselves be seduced by it into accepting Ramsey's basic premise; viz., that the totality of properties already exists before their characterization by definition. Such a conception, I believe, is not far removed from a belief in a platonic realm of ideas which exist in themselves, independently of *if* and *how* finite human beings are able to think them. I think we ought to hold fast to Frege's dictum that, in mathematics, only that may be taken to exist whose existence has been proved (and he meant proved in finitely many steps). I agree with the intuitionists that the finiteness of every logical-mathematical operation, proof, and definition is not required because of some accidental empirical fact about man but is required by the very nature of the subject. Because of this attitude, intuitionist mathematics has been called "anthropological mathematics." It seems to me that, by analogy, we should call Ramsey's mathematics "theological mathematics," for when he speaks of the totality of properties he elevates himself above the actually knowable and definable and in certain respects reasons from the standpoint of an infinite mind which is not bound by the wretched necessity of building every structure step by step.

We may now rephrase our crucial question thus: Can we have Ramsey's result without retaining his absolutist conceptions? His result was this: Limitation to the simple theory of types and retention of the possibility of definitions for mathematical concepts, particularly in real number theory. We can reach this result if, like Ramsey, we allow impredicative definition, but can we do this without falling into his conceptual absolutism? I will try to give an affirmative answer to this question.

Let us go back to the example of the property "inductive" for which we gave an impredicative definition:

$$\mathrm{Ind}(x) =_{\mathrm{Df}} (f)[(\mathrm{Her}(f)\cdot f(0)) \supset f(x)]$$

Let us examine once again whether the use of this definition, i.e., establishing whether the concept holds in an individual case or not, really

leads to circularity and is therefore impossible. According to this definition, that the number two is inductive means:

$$(f)[(\mathrm{Her}(f)\cdot f(0)) \supset f(2)]$$

in words: Every property f which is hereditary and belongs to zero belongs also to two. How can we verify a universal statement of this kind? If we had to examine every single property, an unbreakable circle would indeed result, for then we would run headlong against the property "inductive." Establishing whether something had it would then be impossible in principle, and the concept would therefore be meaningless. But the verification of a universal logical or mathematical sentence does not consist in running through a series of individual cases, for impredicative definitions usually refer to infinite totalities. The belief that we must run through all the individual cases rests on a confusion of "numerical" generality, which refers to objects already given, with "specific" generality (cf. Kaufmann 1930). We do not establish specific generality by running through individual cases but by logically deriving certain properties from certain others. In our example, that the number two is inductive means that the property "belonging to two" follows logically from the property "being hereditary and belonging to zero." In symbols, '$f(2)$' can be derived for an arbitrary f from '$\mathrm{Her}(f)\cdot f(0)$' by logical operations. This is indeed the case. First, the derivation of '$f(0)$' from '$\mathrm{Her}(f)\cdot f(0)$' is trivial and proves the inductiveness of the number zero. The remaining steps are based on the definition of the concept "hereditary":

$$\mathrm{Her}(f) =_{\mathrm{Df}} (n)[f(n) \supset f(n+1)]$$

Using this definition, we can easily show that '$f(0+1)$' and hence '$f(1)$' are derivable from '$\mathrm{Her}(f)\cdot f(0)$' and thereby prove that the number one is inductive. Using this result and our definition, we can derive '$f(1+1)$' and hence '$f(2)$' from '$\mathrm{Her}(f)\cdot f(0)$', thereby showing that the number two is inductive. We see then that the definition of inductiveness, although impredicative, does not hinder its utility. That proofs that the defined property obtains (or does not obtain) in individual cases can be given shows that the definition is meaningful. If we reject the belief that it is necessary to run through individual cases and rather make it clear to ourselves that the complete verification of a statement about an arbitrary property means nothing more than its logical (more exactly, tautological) validity for an arbitrary property, we will come to the conclusion that impredicative definitions are logically admissible. If a property is defined impredicatively, then establishing whether or not it obtains in an individual case may, under certain circumstances, be difficult, or it may even be impossible if there is no solution to the decision problem for that logical system. But in no way does impredicativeness make such decisions impossible in principle for all cases. If the theory just sketched proves feasible, logicism will have been helped over its greatest difficulty, which consists in steering a safe course between the Scylla of the axiom of reducibility and the Charybdis of the allocation of the real numbers to different orders.

Logicism as here described has several features in common both with intuitionism and with formalism. It shares with intuitionism a constructivistic tendency with respect to definition, a tendency which Frege also emphatically endorsed. A concept may not be introduced axiomatically but must be constructed from undefined, primitive concepts step by step through explicit definitions. The admission of impredicative definitions seems at first glance to run counter to this tendency, but this is only true for constructions of the form proposed by Ramsey. Like the intuitionists, we recognize as properties only those expressions (more precisely, expressions of the form of a sentence containing one free variable) which are constructed in finitely many steps from undefined primitive properties of the appropriate domain according to determinate rules of construction. The difference between us lies in the fact that we recognize as valid not only the rules of construction which the intuitionists use (the rules of the so-called "strict functional calculus"), but in addition, permit the use of the expression 'for all properties' (the operations of the so-called "extended functional calculus").

Further, logicism has a methodological affinity with formalism. Logicism proposes to construct the logical-mathematical system in such a way that, although the axioms and rules of inference are chosen with an interpretation of the primitive symbols in mind, nevertheless, *inside the system* the chains of deductions and of definitions are carried through formally as in a pure calculus, i.e., without reference to the meaning of the primitive symbols.

2. The intuitionist foundations of mathematics

[Die intuitionistische Grundlegung der Mathematik]

AREND HEYTING

The intuitionist mathematician proposes to do mathematics as a natural function of his intellect, as a free, vital activity of thought. For him, mathematics is a production of the human mind. He uses language, both

natural and formalized, only for communicating thoughts, i.e., to get others or himself to follow his own mathematical ideas. Such a linguistic accompaniment is not a representation of mathematics; still less is it mathematics itself.

It would be most in keeping with the active attitude of the intuitionist to deal at once with the construction of mathematics. The most important building block of this construction is the concept of unity which is the architectonic principle on which the series of integers depends. The integers must be treated as units which differ from one another only by their place in this series. Since in his *Logischen Grundlagen der exakten Wissenschaften* Natorp has already carried out such an analysis, which in the main conforms tolerably well to the intuitionist way of thinking, I will forego any further analysis of these concepts. But I must still make one remark which is essential for a correct understanding of our intuitionist position: we do not attribute an existence independent of our thought, i.e., a transcendental existence, to the integers or to any other mathematical objects. Even though it might be true that every thought refers to an object conceived to exist independently of it, we can nevertheless let this remain an open question. In any event, such an object need not be completely independent of human thought. Even if they should be independent of individual acts of thought, mathematical objects are by their very nature dependent on human thought. Their existence is guaranteed only insofar as they can be determined by thought. They have properties only insofar as these can be discerned in them by thought. But this possibility of knowledge is revealed to us only by the act of knowing itself. Faith in transcendental existence, unsupported by concepts, must be rejected as a means of mathematical proof. As I will shortly illustrate more fully by an example, this is the reason for doubting the law of excluded middle.

Oskar Becker has dealt thoroughly with the problems of mathematical existence in his book on that subject. He has also uncovered many connections between these questions and the most profound philosophical problems.

We return now to the construction of mathematics. Although the introduction of the fractions as pairs of integers does not lead to any basic difficulties, the definition of the irrational numbers is another story. A real number is defined according to Dedekind by assigning to every rational number either the predicate 'Left' or the predicate 'Right' in such a way that the natural order of the rational numbers is preserved. But if we were to transfer this definition into intuitionist mathematics in exactly this form, we would have no guarantee that Euler's constant C is a real number. We do not need the definition of C. It suffices to know that this definition amounts to an algorithm which permits us to enclose C within an arbitrarily small rational interval. (A rational interval is an interval whose end points are rational numbers. But, as absolutely no ordering relations have been defined between C and the rational numbers, the word 'enclose' is obviously vague for practical purposes. The practical question is that of computing a series of rational intervals each of which is contained in the preceding one in such a way that the computation can always be continued far enough so that the last interval is smaller than an arbitrarily given limit.) But this algorithm still provides us with no way of deciding for an arbitrary rational number A whether it lies left or right of C or is perhaps equal to C. But such a method is just what Dedekind's definition, interpreted intuitionistically, would require.

The usual objection against this argument is that it does not matter whether or not this question can be decided, for, if it is not the case that $A=C$, then either $A<C$ or $A>C$, and this last alternative is decided after a finite, though perhaps unknown, number of steps N in the computation of C. I need only reformulate this objection to refute it. It can mean only this: either there exists a natural number N such that after N steps in the computation of C it turns out that $A<C$ or $A>C$; or there is no such N and hence, of course, $A=C$. But, as we have seen, the existence of N signifies nothing but the possibility of actually producing a number with the requisite property, and the non-existence of N signifies the possibility of deriving a contradiction from this property. Since we do not know whether or not one of these possibilities exists, we may not assert that N either exists or does not exist. In this sense, we can say that the law of excluded middle may not be used here.

In its original form, then, Dedekind's definition cannot be used in intuitionist mathematics. Brouwer, however, has improved it in the following way: Think of the rational numbers enumerated in some way. For the sake of simplicity, we restrict ourselves to the numbers in the closed unit interval and take always as our basis the following enumeration:

$$\text{(A)}\quad 0,\ 1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{2}{3},\ \frac{1}{4},\ \frac{3}{4},\ \frac{1}{5},\ \frac{2}{5},\ \frac{3}{5},\ \frac{4}{5},\ldots$$

A real number is determined by a cut in the series (A); i.e., by a rule which assigns to each rational number in the series either the predicate 'Left' or the predicate 'Right' in such a way that the natural order of the rational numbers is preserved. At each step, however, we permit one individual number to be left out of this mapping. For example, let the rule be so formed that the series of predicates begins this way:

$$0,\ 1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{2}{3},\ \frac{1}{4},\ \frac{3}{4},\ \frac{1}{5},\ \frac{2}{5},\ \frac{3}{5},\ \frac{4}{5},\ldots$$

L, R, L, L, ?, L,

Here 2/3 is temporarily left out of the mapping. We need not know whether or not the predicate for 2/3 is ever determined. But it is also a possibility that 3/4 should become a new excluded number and hence that 2/3 would receive the predicate 'Left'.

It is easy to give a cut for Euler's constant. Let d_n be the smallest difference between two successive numbers in the first n numbers of (A). Now if we compute C far enough to get a rational interval i which is smaller than d_n, then at most one of these n numbers can fall within i. If there is such a number, it becomes the excluded number for the cut. Thus, we can see how closely Brouwer's definition is related to the actual computation of a real number.

We can now take an important step forward. We can drop the requirement that the series of predicates be determined to infinity by a rule. It suffices if the series is determined step by step in some way, e.g., by free choices. I call such sequences "infinitely proceeding." Thus the definition of real numbers is extended to allow infinitely proceeding sequences in addition to rule-determined sequences. Before discussing this new definition in detail, we will give a simple example. We begin with this "Left-Right" choice-sequence:

$$0,\ 1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{2}{3},\ \frac{1}{4},\ \frac{3}{4},\ \frac{1}{5},\ \frac{2}{5},\ \frac{3}{5},\ \frac{4}{5},$$

L, R, L, L, R, L, R, L, L,

Here the question about which predicate 3/5 receives cannot be answered yet, for it must still be decided which predicate to give it. The question about the predicate which 4/5 receives, on the other hand, can be answered now by 'Right,' since that choice would hold for every possible continuation of the sequence. In general, only those questions about an infinitely proceeding sequence which refer to every possible continuation of the sequence are susceptible of a determinate answer. Other questions, like the foregoing about the predicate for 3/5, must therefore be regarded as meaningless. Thus choice-sequences supplant, not so much the individual rule-determined sequences, but rather the totality of all possible rules. A "Left-Right" choice-sequence, the freedom of choice for which is limited only by the conditions which result from the natural order of the rational numbers, determines not just one real number but the spread of all real numbers or the continuum. Whereas we ordinarily think of each real number as individually defined and only afterwards think of them all together, we here define the continuum as a totality. If we restrict this freedom of choice by rules given in advance, we obtain spreads of real numbers. For example, if we prescribe that the sequence begin in the way we have just written it, we define the spread of real numbers between 1/2 and 2/3. An infinitely proceeding sequence gradually becomes a rule-determined sequence when more and more restrictions are placed on the freedom of choice.

We have used the word 'spread' exactly in Brouwer's sense. His definition of a spread is a generalization of this notion. In addition to choice-sequences, Brouwer treats sequences which are formed from choice-sequences by mapping rules. A spread involves two rules. The first rule states which choices of natural numbers are allowed after a determinate finite series of permitted choices has been made. The rule must be so drawn that at least one new permissible choice is known after each finite series of permitted choices has been made. The natural order of the rational numbers is an example of such a rule for our "Left-Right" sequence previously given. The second rule involved in a spread assigns a mathematical object to each permissible choice. The mathematical object may, of course, depend also on choices previously made. Thus it is permissible to terminate the mapping at some particular number and to assign nothing to subsequent choices. A sequence which results from a permissible choice-sequence by a mapping-rule is called an "element" of the spread.

To bring our previous example of the spread of real numbers between 1/2 and 2/3 under this general definition, we will replace the predicates 'Left', 'Right', and 'temporarily undetermined', by 1, 2, and 3; and we will derive the rule for permissible choices from the natural order of the rational numbers and from the requirement that the sequence begin in a particular way; and we will take identity for the mapping-rule.

A spread is not the sum of its elements (this statement is meaningless unless spreads are regarded as existing in themselves). Rather, a spread is identified with its defining rules. Two elements of a spread are said to be equal if equal objects exist at the nth place in both for every n. Equality of elements of a spread, therefore, does not mean that they are the same element. To be the same, they would have to be assigned to the same spread by the same choice-sequence. It would be impractical to call two mathematical objects equal only if they are the same object. Rather, every kind of object must receive its own definition of equality.

Brouwer calls "species" those spreads which are defined, in classical terminology, by a characteristic property of their members. A species,

like a spread, is not regarded as the sum of its members but is rather identified with its defining property. Impredicative definitions are made impossible by the fact, which intuitionists consider self-evident, that only previously defined objects may occur as members of a species. There results, consequently, a step-by-step introduction of species. The first level is made up of those spread-species whose defining property is identity with an element of a particular spread. Hence, to every spread M there corresponds the spread-species of those spread-elements which are identical with some element of M.[1] A species of the first order can contain spread-elements and spread-species. In addition, a species of the second order contains species of the first order as members, and so on.

The introduction of infinitely proceeding sequences is not a necessary consequence of the intuitionist approach. Intuitionist mathematics could be constructed without choice-sequences. But the following set-theoretic theorem about the continuum shows how much mathematics would thereby be impoverished. This theorem will also serve as an example of an intuitionist reasoning process.

Let there be a rule assigning to each real number a natural number as its correlate. Assume that the real numbers a and b have different correlates, e.g., 1 and 2. Then, by a simple construction, we can determine a third number c which has the following property: in every neighborhood of c, no matter how small, there is a mapped number other than c; i.e., every finite initial segment of the cut which defines c can be continued so as to get a mapped number other than c. We define the number d by a choice-sequence thus: we begin as with c but we reserve the freedom to continue at an arbitrary choice in a way different from that for c. Obviously the correlate of d is not determined after any previously known finite number of choices. Accordingly, no definite correlate is assigned to d. But this conclusion contradicts our premise that every real number has a correlate. Our assumption that the two numbers a and b have different correlates is thus shown to be contradictory. And, since two natural numbers which cannot be distinguished are the same number, we have the following theorem: if every real number is assigned a correlate, then all the real numbers have the same correlate.

As a special result, we have: if a continuum is divided into two subspecies in such a way that every member belongs to one and only one of these subspecies, then one of the subspecies is empty and other other is identical with the continuum.

The unit continuum, for example, cannot be subdivided into the species of numbers between 0 and 1/2 and the species of numbers between

[1]This definition of spread-species is taken from a communication of Professor Brouwer.

1/2 and 1, for the preceding construction produces a number for which one need never decide whether it is larger or smaller than 1/2. The theorems about the continuity of a function determined in an interval are also connected with the foregoing theorem. But Brouwer's theorem about the uniform continuity of all full functions goes far beyond these results.

But what becomes of the theorem we have just proved if no infinitely proceeding sequences are allowed in mathematics? In that event, the species of numbers defined by rule-determined sequences would have to take the place of the continuum. This definition is admissible if we take it to mean that a number belongs to this species only if there is a rule which permits us actually to determine all the predicates of the sequence successively.

In this event, the foregoing proof continues to hold only if we succeed in defining the number d by a rule-determined sequence rather than by a choice-sequence. We can probably do it if we make use of certain unresolved problems; e.g., whether or not the sequence 0123456789 occurs in the decimal expansion of π. We can let the question – whether or not to deviate from the predicate series for c, at the nth predicate in the predicate sequence for d – depend on the occurrence of the preceding sequence at the nth digit after the decimal point in π. This proof obviously is weakened as soon as the question about the sequence is answered. But, in the event that it is answered, we can replace this question by a similar unanswered question, if there are any left. We can prove our theorem for rule-determined sequences only on the condition that there always remain unsolved problems. More precisely, the theorem is true if there are two numbers, determined by rule-determined sequences, such that the question about whether they are the same or different poses a demonstrably unsolvable problem. It is false if the existence of two such numbers is contradictory. But the problem raised by these conditions is insuperable. Even here choice-sequences prove to be superior to rule-determined sequences in that the former make mathematics independent of the question of the existence of unsolvable problems.

We conclude our treatment of the construction of mathematics in order to say something about the intuitionist propositional calculus. We here distinguish between propositions and assertions. An assertion is the affirmation of a proposition. A mathematical proposition expresses a certain expectation. For example, the proposition, 'Euler's constant C is rational', expresses the expectation that we could find two integers a and b such that $C=a/b$. Perhaps the word 'intention', coined by the phenomenologists, expresses even better what is meant here. We also use the word 'proposition' for the intention which is linguistically expressed by

the proposition. The intention, as already emphasized above, refers not only to a state of affairs thought to exist independently of us but also to an experience thought to be possible, as the preceding example clearly brings out.

The affirmation of a proposition means the fulfillment of an intention. The assertion 'C is rational', for example, would mean that one has in fact found the desired integers. We distinguish an assertion from its corresponding proposition by the assertion sign '$\vdash$' that Frege introduced and which Russell and Whitehead also used for this purpose. The affirmation of a proposition is not itself a proposition; it is the determination of an empirical fact, viz., the fulfillment of the intention expressed by the proposition.

A logical function is a process for forming another proposition from a given proposition. Negation is such a function. Becker, following Husserl, has described its meaning very clearly. For him negation is something thoroughly positive, viz., the intention of a contradiction contained in the original intention. The proposition 'C is not rational', therefore, signifies the expectation that one can derive a contradiction from the assumption that C is rational. It is important to note that the negation of a proposition always refers to a proof procedure which leads to the contradiction, even if the original proposition mentions no proof procedure. We use $\neg$ as the symbol for negation.

For the law of excluded middle we need the logical function "either-or". '$p \vee q$' signifies that intention which is fulfilled if and only if at least one of the intentions p and q is fulfilled. The formula for the law of excluded middle would be '$\vdash p \vee \neg p$'. One can assert this law for a particular proposition p only if p either has been proved or reduced to a contradiction. Thus, a proof that the law of excluded middle is a general law must consist in giving a method by which, when given an arbitrary proposition, one could always prove either the proposition itself or its negation. Thus the formula '$p \vee \neg p$' signifies the expectation of a mathematical construction (method of proof) which satisfies the aforementioned requirement. Or, in other words, this formula is a mathematical proposition; the question of its validity is a mathematical problem which, when the law is stated generally, is unsolvable by mathematical means. In this sense, logic is dependent on mathematics.

We conclude with some remarks on the question of the solvability of mathematical problems. A problem is posed by an intention whose fulfillment is sought. It is solved either if the intention is fulfilled by a construction or if it is proved that the intention leads to a contradiction. The question of solvability can, therefore, be reduced to that of provability.

A proof of a proposition is a mathematical construction which can itself be treated mathematically. The intention of such a proof thus yields a new proposition. If we symbolize the proposition 'the proposition p is provable' by '$+p$', then '$+$' is a logical function, viz., "provability." The assertions '$\vdash p$' and '$\vdash +p$' have exactly the same meaning. For, if p is proved, the provability of p is also proved, and if $+p$ is proved, then the intention of a proof of p has been fulfilled, i.e., p has been proved. Nevertheless, the propositions p and $+p$ are not identical, as can best be made clear by an example. In the computation of Euler's constant C, it can happen that a particular rational value, say A, is contained for an unusually long time within the interval within which we keep more narrowly enclosing C so that we finally suspect that $C=A$; i.e., we expect that, if we continued the computation of C, we would keep on finding A within this interval. But such a suspicion is by no means a proof that it will always happen. The proposition $+(C=A)$, therefore, contains more than the proposition $(C=A)$.

If we apply negation to both of these propositions, then we get not only two different propositions, '$\neg p$' and '$\neg +p$', but also the assertions, '$\vdash \neg p$' and '$\vdash \neg +p$', are different. '$\vdash \neg +p$' means that the assumption of such a construction as $+p$ requires is contradictory. The simple expectation p, however, need not lead to a contradiction. Here is how this works in our example just cited. Assume that we have proved the contradictoriness of the assumption that there is a construction which proves that A lies within every interval that contains C ($\vdash \neg +p$). But still the assumption that in the actual computation of C we will always in fact find A within our interval need not lead to a contradiction. It is even conceivable that we might prove that the latter assumption could never be proved to be contradictory, and hence that we could assert at the same time both '$\vdash \neg +p$' and '$\vdash \neg \neg p$'. In such an event, the problem whether $C=A$ would be essentially unsolvable.

The distinction between p and $+p$ vanishes as soon as a construction is intended in p itself, for the possibility of a construction can be proved only by its actual execution. If we limit ourselves to those propositions which require a construction, the logical function of provability generally does not arise. We can impose this restriction by treating only propositions of the form 'p is provable' or, to put it another way, by regarding every intention as having the intention of a construction for its fulfillment added to it. It is in this sense that intuitionist logic, insofar as it has been developed up to now without using the function $+$, must be understood. The introduction of provability would lead to serious complications. Yet its minimal practical value would hardly make it worthwhile to deal with those complications in detail.[2] But here this notion has given us an insight into how to conceive of essentially unsolvable problems.

We will have accomplished our purpose if we have shown you that intuitionism contains no arbitrary assumptions. Still less does it contain artificial prohibitions, such as those used to avoid the logical paradoxes. Rather, once its basic attitude has been adopted, intuitionism is the only possible way to construct mathematics.

3. The formalist foundations of mathematics

JOHANN VON NEUMANN

I

Critical studies of the foundations of mathematics during the past few decades, in particular Brouwer's system of "intuitionism," have reopened the question of the origins of the generally supposed absolute validity of classical mathematics. Noteworthy is the fact that this question, in and of itself philosophico-epistemological, is turning into a logico-mathematical one. As a result of three important advances in the field of mathematical logic (namely: Brouwer's sharp formulation of the defects of classical mathematics; Russell's thorough and exact description of its methods – both the good and the bad; and Hilbert's contributions to the mathematical-combinatorial investigation of these methods and their relations), more and more it is unambiguous mathematical questions, not matters of taste, that are being investigated in the foundation of mathematics. As the other papers have dealt extensively both with the domain (delimited by Brouwer) of unconditionally valid (i.e., needing no justification) "intuitionist" or "finitistic" definitions and methods of proof and with Russell's formal characterization (which has been further developed by his school) of the nature of classical mathematics, we need not dwell on these topics any longer. An understanding of them is, of course, a necessary prerequisite for an understanding of the utility, tendency, and *modus procedendi* of Hilbert's theory of proof. We turn instead directly to the theory of proof.

The leading idea of Hilbert's theory of proof is that, even if the statements of classical mathematics should turn out to be false as to content, nevertheless, classical mathematics involves an internally closed procedure which operates according to fixed rules known to all mathematicians

[2]The question dealt with in this paragraph was fully clarified only in a discussion with H. Freudenthal after the conference. The results of this discussion are reproduced in the text.

and which consists basically in constructing successively certain combinations of primitive symbols which are considered "correct" or "proved." This construction-procedure, moreover, is "finitary" and directly constructive. To see clearly the essential difference between the occasionally non-constructive handling of the "content" of mathematics (real numbers and the like) and the always constructive linking of the steps in a proof, consider this example: Assume that there exists a classical proof of the existence of a real number x with a certain very complicated and deep-seated property $E(x)$. Then it may well happen that, from this proof, we can in no way derive a procedure for constructing an x such that $E(x)$. (We shall give an example of such a proof in a moment.) On the other hand, if the proof somehow violated the conventions of mathematical inference, i.e., if it contained an error, we could, of course, find this error by a finitary process of checking. In other words, although the content of a classical mathematical sentence cannot always (i.e., generally) be finitely verified, the formal way in which we arrive at the sentence can be. Consequently, if we wish to prove the validity of classical mathematics, which is possible in principle only by reducing it to the *a priori* valid finitistic system (i.e., Brouwer's system), then we should investigate, not statements, but methods of proof. We must regard classical mathematics as a combinatorial game played with the primitive symbols, and we must determine in a finitary combinatorial way to which combinations of primitive symbols the construction methods or "proofs" lead.

As we promised, we now produce an example of a non-constructive existence proof. Let $f(x)$ be a function which is linear from 0 to 1/3, from 1/3 to 2/3, from 2/3 to 1, and so on. Let

$$f(0)=-1;\quad f\left(\frac{1}{3}\right)=-\sum_{n=1}^{n=\infty}\frac{\epsilon_{2n}}{2^n};\quad f\left(\frac{2}{3}\right)=\sum_{n=1}^{n=\infty}\frac{\epsilon_{2n}}{2^n};\quad \text{and}\quad f(1)=1$$

ϵ_n is defined as follows: if $2k$ is the sum of two prime numbers, then $\epsilon_k=0$; otherwise $\epsilon_k=1$. Obviously $f(x)$ is continuous and calculable with arbitrary accuracy at any point x. Since $f(0)<0$ and $f(1)>0$, there exists an x, where $0\leqslant x\leqslant 1$, such that $f(x)=0$. (In fact we readily see that $1/3\leqslant x\leqslant 2/3$.) However the task of finding a root with an accuracy greater than $\pm 1/6$ encounters formidable difficulties. Given the present state of mathematics, these difficulties are insuperable, for if we could find such a root, then we could predict with certitude the existence of a root $<2/3$ or $>1/3$, according as its approximate value were $\leqslant 1/2$ or $\geqslant 1/2$, respectively. The former case (where the approximate value of the root is $\leqslant 1/2$) excludes both that $f(1/3)<0$ and that $f(2/3)=0$; the latter case (where the approximate value of the root $\geqslant 1/2$) excludes both that $f(1/3)=0$ and that $f(2/3)>0$. In other words, in the former case the value of ϵ_n must be 0 for all even n but not for all odd n; in the latter case the value of ϵ_n must be 0 for all odd n but not for all even n. Hence we would have proved that Goldbach's famous conjecture (that $2n$ is always the sum of two prime numbers), instead of holding universally, must already fail to hold for odd n in the former case and for even n in the latter. But no mathematician today can supply a proof for either case, since no one can find the solution of $f(x)=0$ more accurately than with an error of 1/6. (With an error of 1/6, 1/2 would be an approximate value of the root, for the root lies between 1/3 and 2/3, i.e., between $1/2-1/6$ and $1/2+1/6$.)

II

Accordingly, the tasks which Hilbert's theory of proof must accomplish are these:

1. To enumerate all the symbols used in mathematics and logic. These symbols, called "primitive symbols," include the symbols '$\sim$' and '$\rightarrow$' (which stand for "negation" and "implication" respectively).
2. To characterize unambiguously all the combinations of these symbols which represent statements classified as "meaningful" in classical mathematics. These combinations are called "formulas." (Note that we said only "meaningful," not necessarily "true." '$1+1=2$' is meaningful but so is '$1+1=1$', independently of the fact that one is true and the other false. On the other hand, combinations like '$1+\rightarrow=1$' and '$++1=\rightarrow$' are meaningless.)
3. To supply a construction procedure which enables us to construct successively all the formulas which correspond to the "provable" statements of classical mathematics. This procedure, accordingly, is called "proving."
4. To show (in a finitary combinatorial way) that those formulas which correspond to statements of classical mathematics which can be checked by finitary arithmetical methods can be proved (i.e., constructed) by the process described in (3) if and only if the check of the corresponding statement shows it to be true.

To accomplish tasks 1–4 would be to establish the validity of classical mathematics as a short-cut method for validating arithmetical statements whose elementary validation would be much too tedious. But since this is in fact the way we use mathematics, we would at the same time sufficiently establish the empirical validity of classical mathematics.

We should remark that Russell and his school have almost completely accomplished tasks 1–3. In fact, the formalization of logic and mathematics suggested by tasks 1–3 can be carried out in many different ways. The real problem, then, is (4).

In connection with (4) we should note the following: If the "effective check" of a numerical formula shows it to be false, then from that formula we can derive a relation $p=q$ where p and q are two different, effectively given numbers. Hence (according to task 3) this would give us a formal proof of $p=q$ from which we could obviously get a proof of $1=2$. Therefore, the sole thing we must show to establish (4) is the formal unprovability of $1=2$; i.e., we need to investigate only this one particular false numerical relation. The unprovability of the formula $1=2$ by the methods described in (3) is called "consistency." The real problem, then, is that of finding a finitary combinatorial proof of consistency.

III

To be able to indicate the direction which a proof of consistency takes, we must consider formal proof procedure – as in (3) – a little more closely. It is defined as follows:

3_1. Certain formulas, characterized in an unambiguous and finitary way, are called "axioms." Every axiom is considered proved.

3_2. If a and b are two meaningful formulas, and if a and $a \rightarrow b$ have both been proved, then b also has been proved.

Note that, although (3_1) and (3_2) do indeed enable us to write down successively all provable formulas, still this process can never be finished. Further, (3_1) and (3_2) contain no procedure for deciding whether a given formula e is provable. As we cannot tell in advance which formulas must be proved successively in order ultimately to prove e, some of them might turn out to be far more complicated and structurally quite different from e itself. (Anyone who is acquainted, for example, with analytic number theory knows just how likely this possibility is, especially in the most interesting parts of mathematics.) But the problem of deciding the provability of an arbitrarily given formula by means of a (naturally finitary) general procedure, i.e., the so-called decision problem for mathematics, is much more difficult and complex than the problem discussed here.

As it would take us too far afield to give the axioms which are used in classical mathematics, the following remarks must suffice to characterize them. Although infinitely many formulas are regarded as axioms (for example, by our definition each of the formulas $1=1$, $2=2$, $3=3, \ldots$ is

an axiom), they are nevertheless constructed from finitely many schemata by substitution in this manner: 'If a, b, and c are formulas, then $(a \rightarrow b) \rightarrow ((b \rightarrow c) \rightarrow (a \rightarrow c))$ is an axiom', and the like.

Now if we could succeed in producing a class R of formulas such that

(α) Every axiom belongs to R,
(β) If a and $a \rightarrow b$ belong to R, then b also belongs to R,
(γ) '$1=2$' does not belong to R,

then we would have proved consistency, for according to (α) and (β) every proved formula obviously must belong to R, and according to (γ), $1=2$ must therefore be unprovable. The actual production of such a class at this time is unthinkable, however, for it poses difficulties comparable to those raised by the decision problem. But the following remark leads from this problem to a much simpler one: If our system were inconsistent, then there would exist a proof of $1=2$ in which only a finite number of axioms are used. Let the set of these axioms be called M. Then the axiom system M is already inconsistent. Hence the axiom system of classical mathematics is certainly consistent if every finite subsystem thereof is consistent. And this is surely the case if, for every finite set of axioms M, we can give a class of formulas R_M which has the following properties:

(α) Every axiom of M belongs to R_M.
(β) If a and $a \rightarrow b$ belong to R_M, then b also belongs to R_M.
(γ) $1=2$ does not belong to R_M.

This problem is not connected with the (much too difficult) decision problem, for R_M depends only on M and plainly says nothing about provability (with the help of all the axioms). It goes without saying that we must have an effective, finitary procedure for constructing R_M (for every effectively given finite set of axioms M) and that the proofs of (α), (β), and (γ) must also be finitary.

Although the consistency of classical mathematics has not yet been proved, such a proof has been found for a somewhat narrower mathematical system. This system is closely related to a system which Weyl proposed before the conception of the intuitionist system. It is substantially more extensive than the intuitionist system but narrower than classical mathematics (for bibliographical material, see Weyl 1927).

Thus Hilbert's system has passed the first test of strength: the validity of a non-finitary, not purely constructive mathematical system has been established through finitary constructive means. Whether someone will succeed in extending this validation to the more difficult and more important system of classical mathematics, only the future will tell.

Disputation

AREND HEYTING

Persons of the dialogue: *Class, Form, Int, Letter, Prag, Sign*

Class: How do you do, Mr. Int? Did you not flee the town on this fine summer day?

Int: I had some ideas and worked them out at the library.

Class: Industrious bee! How are you getting along?

Int: Quite well. Shall we have a drink?

Class: Thank you. I bet you worked on that hobby of yours, rejection of the excluded middle, and the rest. I never understood why logic should be reliable everywhere else, but not in mathematics.

Int: We have spoken about that subject before. The idea that for the description of some kinds of objects another logic may be more adequate than the customary one has sometimes been discussed. But it was Brouwer who first discovered an object which actually requires a different form of logic, namely the mental mathematical construction [L. E. J. Brouwer 1908]. The reason is that in mathematics from the very beginning we deal with the infinite, whereas ordinary logic is made for reasoning about finite collections.

Class: I know, but in my eyes logic is universal and applies to the infinite as well as to the finite.

Int: You ought to consider what Brouwer's program was [L. E. J. Brouwer 1907]. It consisted in the investigation of mental mathematical construction as such, without reference to questions regarding the nature of the constructed objects, such as whether these objects exist independently of our knowledge of them. That this point of view leads immediately to the rejection of the principle of excluded middle, I can best demonstrate by an example.

Let us compare two definitions of natural numbers, say k and l.

Excerpted by kind permission of the author and publisher from *Intuitionism: an Introduction*, Arend Heyting, North-Holland, 1956 (3rd ed., 1971).

I. k is the greatest prime such that $k-1$ is also a prime, or $k=1$ if such a number does not exist.

II. l is the greatest prime such that $l-2$ is also a prime, or $l=1$ if such a number does not exist.

Classical mathematics neglects altogether the obvious difference in character between these definitions. k can actually be calculated ($k=3$), whereas we possess no method for calculating l, as it is not known whether the sequence of pairs of twin primes $p, p+2$ is finite or not. Therefore intuitionists reject II as a definition of an integer; they consider an integer to be well-defined only if a method for calculating it is given. Now this line of thought leads to the rejection of the principle of excluded middle, for if the sequence of twin primes were either finite or not finite, II would define an integer.

Class: One may object that the extent of our knowledge about the existence or non-existence of a last pair of twin primes is purely contingent and entirely irrelevant in questions of mathematical truth. Either an infinity of such pairs exist, in which case $l=1$; or their number is finite, in which case l equals the greatest prime such that $l-2$ is also a prime. In every conceivable case l is defined; what does it matter whether or not we can actually calculate the number?

Int: Your argument is metaphysical in nature. If "to exist" does not mean "to be constructed", it must have some metaphysical meaning. It cannot be the task of mathematics to investigate this meaning or to decide whether it is tenable or not. We have no objection against a mathematician privately admitting any metaphysical theory he likes, but Brouwer's program entails that we study mathematics as something simpler, more immediate than metaphysics. In the study of mental mathematical constructions "to exist" must be synonymous with "to be constructed".

Class: That is to say, as long as we do not know if there exists a last pair of twin primes, II is not a definition of an integer, but as soon as this problem is solved, it suddenly becomes such a definition. Suppose on January 1, 1970 it is proved that an infinity of twin primes exists; from that moment $l=1$. Was $l=1$ before that date or not? [Menger 1930].

Int: A mathematical assertion affirms the fact that a certain mathematical construction has been effected. It is clear that before the construction was made, it had not been made. Applying this remark to your example, we see that before Jan. 1, 1970 it had not been proved that $l=1$. But this is not what you mean. It seems to me that in order to clarify the sense of your question you must again refer to metaphysical concepts: to some world of mathematical things existing independently of our knowledge, where "$l=1$" is true in some absolute sense. But I repeat that mathematics ought not to depend upon such notions as these. In fact all mathematicians and even intuitionists are convinced that in some sense mathematics bear upon eternal truths, but when trying to define precisely this sense, one gets entangled in a maze of metaphysical difficulties. The only way to avoid them is to banish them from mathematics. This is what I meant by saying that we study mathematical constructions as such and that for this study classical logic is inadequate.

Class: Here come our friends Form and Letter. Boys, we are having a most interesting discussion on intuitionism.

Letter: Could you speak about anything else with good old Int? He is completely submerged in it.

Int: Once you have been struck with the beauty of a subject, devote your life to it!

Form: Quite so! Only I wonder how there can be beauty in so indefinite a thing as intuitionism. None of your terms are well-defined, nor do you give exact rules of derivation. Thus one for ever remains in doubt as to which reasonings are correct and which are not [R. Carnap 1934b, p. 41; 1937, p. 46; W. Dubislav 1932, pp. 57, 75]. In daily speech no word has a perfectly fixed meaning; there is always some amount of free play, the greater, the more abstract the notion is. This makes people miss each other's point, also in non-formalized mathematical reasonings. The only way to achieve absolute rigour is to abstract all meaning from the mathematical statements and to consider them for their own sake, as sequences of signs, neglecting the sense they may convey. Then it is possible to formulate definite rules for deducing new statements from those already known and to avoid the uncertainty resulting from the ambiguity of language.

Int: I see the difference between formalists and intuitionists mainly as one of taste. You also use meaningful reasoning in what Hilbert called metamathematics, but your purpose is to separate these reasonings from purely formal mathematics and to confine yourself to the most simple reasonings possible. We, on the contrary, are interested not in the formal side of mathematics, but exactly in that type of reasoning which appears in metamathematics; we try to develop it to its farthest consequences. This preference arises from the conviction that we find here one of the most fundamental faculties of the human mind.

Form: If you will not quarrel with formalism, neither will I with intuitionism. Formalists are among the most pacific of mankind. Any theory

may be formalized and then becomes subject to our methods. Also intuitionistic mathematics may and will be thus treated [R. Carnap 1934b, p. 44; 1937, p. 51].

Class: That is to say, intuitionistic mathematics ought to be studied as a part of mathematics. In mathematics we investigate the consequences of given assumptions; the intuitionistic assumptions may be interesting, but they have no right to a monopoly.

Int: Nor do we claim that; we are content if you admit the good right of our conception. But I must protest against the assertion that intuitionism starts from definite, more or less arbitrary assumptions. Its subject, constructive mathematical thought, determines uniquely its premises and places it beside, not interior to, classical mathematics, which studies another subject, whatever subject that may be. For this reason an agreement between formalism and intuitionism by means of the formalization of intuitionistic mathematics is also impossible. It is true that even in intuitionistic mathematics the finished part of a theory may be formalized. It will be useful to reflect for a moment upon the meaning of such a formalization. We may consider the formal system as the linguistic expression, in a particularly suitable language, of mathematical thought.

If we adopt this point of view, we clash against the obstacle of the fundamental ambiguousness of language. As the meaning of a word can never be fixed precisely enough to exclude every possibility of misunderstanding, we can never be mathematically sure that the formal system expresses correctly our mathematical thoughts.

However, let us take another point of view. We may consider the formal system itself as an extremely simple mathematical structure; its entities (the signs of the system) are associated with other, often very complicated, mathematical structures. In this way formalizations may be carried out inside mathematics, and it becomes a powerful mathematical tool. Of course, one is never sure that the formal system represents fully any domain of mathematical thought; at any moment the discovering of new methods of reasoning may force us to extend the formal system.

Form: For several years we have been familiar with this situation. Gödel's incompleteness theorem showed us that any consistent formal system of number-theory may be extended consistently in different ways.

Int: The difference is that intuitionism proceeds independently of the formalization, which can but follow after the mathematical construction.

Class: What puzzles me most is that you both seem to start from nothing at all. You seem to be building castles in the air. How can you know if your reasoning is sound if you do not have at your disposal the infallible criterion given by logic? Yesterday I talked with Sign, who is still more of a relativist than either of you. He is so slippery that no argument gets hold of him, and he never comes to any somewhat solid conclusion. I fear this fate for anybody who discards the support of logic, that is, of common sense.

Sign: Speak of the devil and his imp appears. Were you speaking ill of me?

Class: I alluded to yesterday's discussion. To-day I am attacking these other two damned relativists.

Sign: I should like to join you in that job, but first let us hear the reply of your opponents. Please meet my friend Prag; he will be interested in the discussion.

Form: How do you do? Are you also a philosopher of science?

Prag: I hate metaphysics.

Int: Welcome, brother!

Form: Why, I would rather not defend my own position at the moment, as our discussion has dealt mainly with intuitionism and we might easily confuse it. But I fear that you are wrong as to intuitionistic logic. It has indeed been formalized and valuable work in this field has been done by a score of authors. This seems to prove that intuitionists esteem logic more highly than you think, though it is another logic than you are accustomed to.

Int: I regret to disappoint you. Logic is not the ground upon which I stand. How could it be? It would in turn need a foundation, which would involve principles much more intricate and less direct than those of mathematics itself. A mathematical construction ought to be so immediate to the mind and its result so clear that it needs no foundation whatsoever. One may very well know whether a reasoning is sound without using any logic; a clear scientific conscience suffices. Yet it is true that intuitionistic logic has been developed. To indicate what its significance is, let me give you an illustration. Let A designate the property of an integer of being divisible by 8, B the same by 4, C the same by 2. For $8a$ we may write $4\times 2a$; by this mathematical construction P we see that the property A entails B ($A\rightarrow B$). A similar construction Q shows $B\rightarrow C$. By effecting first P, then Q (juxtaposition of P and Q) we obtain $8a=2\times(2\times 2a)$ showing $A\rightarrow C$. This process remains valid if for A, B, C we substitute arbitrary properties: If the construction P shows that $A\rightarrow B$ and Q shows that $B\rightarrow C$, then the juxtaposition of P and Q shows that $A\rightarrow C$. We have obtained a logical theorem. The process by which it is deduced shows us that it does not differ essentially from mathematical theorems; it is only more general, e.g., in the same sense that "addition of integers

is commutative" is a more general statement than "$2+3=3+2$". This is the case for every logical theorem: it is but a mathematical theorem of extreme generality; that is to say, logic is a part of mathematics, and can by no means serve as a foundation for it. At least, this is the conception of logic to which I am naturally led; it may be possible and desirable to develop other forms of logic for other purposes.

It is the mathematical logic which I just described that has been formalized. The resulting formal system proves to have peculiar properties, very interesting when compared to those of other systems of formal logic. This fact has led to the investigations to which Mr. Form alluded, but, however interesting, they are tied but very loosely to intuitionistic mathematics.

Letter: In my opinion all these difficulties are imaginary or artificial. Mathematics is quite a simple thing. I define some signs and I give some rules for combining them; that is all.

Form: You want some modes of reasoning to prove the consistency of your formal system.

Letter: Why should I want to prove it? You must not forget that our formal systems are constructed with the aim towards applications and that in general they prove useful; this fact would be difficult to explain if every formula were deducible in them. Thereby we get a practical conviction of consistency which suffices for our work. What I contest in intuitionism is the opinion that mathematics has anything to do with the infinite. I can write down a sign, say α, and call it the cardinal number of the integers. After that I can fix rules for its manipulation in agreement with those which Mr. Class uses for this notion; but in doing this I operate entirely in the finite. As soon as the notion of infinity plays a part, obscurity and confusion penetrate into the reasoning. Thus all the intuitionistic assertions about the infinite seem to me highly ambiguous, and it is even questionable whether such a sign as $10^{10^{10}}$ has any other meaning than as a figure on paper with which we operate according to certain rules [J. Dieudonné 1951].

Int: Of course your extreme finitism grants the maximum of security against misunderstanding, but in our eyes it implies a denial of understanding which it is difficult to accept. Children in the elementary school understand what the natural numbers are and they accept the fact that the sequence of natural numbers can be indefinitely continued.

Letter: It is suggested to them that they understand.

Int: That is no objection, for every communication by means of language may be interpreted as suggestion. Also Euclid in the 20th proposition of Book IX, where he proved that the set of prime numbers is infinite, knew what he spoke about. This elementary notion of natural numbers, familiar to every thinking creature, is fundamental in intuitionistic mathematics. We do not claim for it any form of certainty or definiteness in an absolute sense, which would be unrealizable, but we contend that it is sufficiently clear to build mathematics upon.

Letter: My objection is that you do not suppose too little, as Mr. Class thinks, but far too much. You start from certain principles which you take as intuitively clear without any explanation and you reject other modes of reasoning without giving any grounds for that discrimination. For instance, to most people the principle of the excluded middle seems at least as evident as that of complete induction. Why do you reject the former and accept the latter? Such an unmotivated choice of first principles gives to your system a strongly dogmatic character.

Int: Indeed intuitionistic assertions must seem dogmatic to those who read them as assertions about facts, but they are not meant in this sense. Intuitionistic mathematics consists, as I have explained already to Mr. Class, in mental constructions; a mathematical theorem expresses a purely empirical fact, namely the success of a certain construction. "2+2=3+1" must be read as an abbreviation for the statement: "I have effected the mental constructions indicated by "2+2" and by "3+1" and I have found that they lead to the same result." Now tell me where the dogmatic element can come in; not in the mental construction itself, as is clear by its very nature as an activity, but no more in the statements made about the constructions, for they express purely empirical results.

Letter: Yet you contend that these mental constructions lead to some sort of truth; they are not a game of solitaire, but in some sense must be of value for mankind, or you would be wrong in annoying others with them. It is in this pretence that I see the dogmatic element. The mathematical intuition inspires you with objective and eternal truths; in this sense your point of view is not only dogmatic, but even theological [H. B. Curry 1951, p. 6].

Int: In the first instance, my mathematical thoughts belong to my individual intellectual life and are confined to my personal mind, as is the case for other thoughts as well. We are generally convinced that other people have thoughts analogous to our own and that they can understand us when we express our thoughts in words, but we also know that we are never quite sure of being faultlessly understood. In this respect, mathematics does not essentially differ from other subjects; if for this reason you consider mathematics to be dogmatic, you ought to call any human reasoning dogmatic. The characteristic of mathematical thought is, that

it does not convey truth about the external world, but is only concerned with mental constructions. Now we must distinguish between the simple practice of mathematics and its valuation. In order to construct mathematical theories no philosophical preliminaries are needed, but the value we attribute to this activity will depend upon our philosophical ideas.

Sign: In the way you treat language you put the clock back. Primitive language has this floating, unsteady character you describe, and the language of daily life is still in the main of the same sort, but as soon as scientific thought begins, the formalization of language sets in. In the last decades significists have studied this process. It has not yet come to an end, for more strictly formalized languages are still being formed.

Int: If really the formalization of language is the trend of science, then intuitionistic mathematics does not belong to science in this sense of the word. It is rather a phenomenon of life, a natural activity of man, which itself is open to study by scientific methods; it has actually been studied by such methods, namely that of formalizing intuitionistic reasoning and the signific method, but it is obvious that this study does not belong to intuitionistic mathematics, nor do its results. That such a scientific examination of intuitionistic mathematics will never produce a complete and definite description of it, no more than a complete theory of other phenomena is attainable, is clearly to be seen. Helpful and interesting as these metaintuitionistic considerations may be, they cannot be incorporated into intuitionistic mathematics itself. Of course, these remarks do not apply to formalization inside mathematics, as I described it a few moments ago.

Prag: Allow me to underline what Mr. Sign said just now. Science proceeds by formalization of language; it uses this method because it is efficient. In particular the modern completely formalized languages have appeared to be most useful. The ideal of the modern scientist is to prepare an arsenal of formal systems ready for use from which he can choose, for any theory, that system which correctly represents the experimental results. Formal systems ought to be judged by this criterion of usefulness and not by a vague and arbitrary interpretation, which is preferred for dogmatic or metaphysical reasons.

Int: It seems quite reasonable to judge a mathematical system by its usefulness. I admit that from this point of view intuitionism has as yet little chance of being accepted, for it would be premature to stress the few weak indications that it might be of some use in physics [J. L. Destouches 1951]; in my eyes its chances of being useful for philosophy, history and the social sciences are better. In fact, mathematics, from the intuitionistic point of view, is a study of certain functions of the human mind, and as such it is akin to these sciences. But is usefulness really the only measure of value? It is easy to mention of score of valuable activities Ua which in no way support science, such as the arts, sports, and light entertainment. We claim for intuitionism a value of this sort, which it is difficult to define beforehand, but which is clearly felt in dealing with the matter. You know how philosophers struggle with the problem of defining the concept of value in art; yet every educated person feels this value. The case is analogous for the value of intuitionistic mathematics.

Form: For most mathematicians this value is affected fatally by the fact that you destroy the most precious mathematical results; a valuable method for the foundation of mathematics ought to save as much as possible of its results [D. Hilbert 1922]. This might even succeed by constructive methods; for definitions of constructiveness other than that advocated by the intuitionists are conceivable. For that matter, even the small number of actual intuitionists do not completely agree about the delimitation of the constructive. The most striking example is the rejection by Griss of the notion of negation, which other intuitionists accept as perfectly clear [H. Freudenthal 1936; G. F. C. Griss 1946a, p. 24; 1946b]. It seems probable, on the other hand, that a somewhat more liberal conception of the constructive might lead to the saving of the vital parts of classical mathematics.

Int: As intuitionists speak a non-formalized language, slight divergences of opinion between them can be expected. Though they have arisen sooner and in more acute forms than we could foresee, they are in no way alarming, for they all concern minor points and do not affect the fundamental ideas, about which there is complete agreement. Thus it is most unlikely that a wider conception of constructiveness could obtain the support of intuitionists. As to the mutilation of mathematics of which you accuse me, it must be taken as an inevitable consequence of our standpoint. It can also be seen as the excision of noxious ornaments, beautiful in form, but hollow in substance, and it is at least partly compensated for by the charm of subtle distinctions and witty methods by which intuitionists have enriched mathematical thought.

Form: Our discussion has assumed the form of a discussion of values. I gather from your words that you are ready to acknowledge the value of other conceptions of mathematics, but that you claim for your conception a value of its own. Is that right?

Int: Indeed, the only positive contention in the foundation of mathematics which I oppose is that classical mathematics has a clear sense; I must confess that I do not understand that. But even those who maintain

that they do understand it might still be able to grasp our point of view and to value our work.

Letter: It is shown by the paradoxes that classical mathematics is not perfectly clear.

Form: Yes, but intuitionistic criticism goes much farther than is necessary to avoid the paradoxes; Mr. Int has not even mentioned them as an argument for his conception, and no doubt in his eyes consistency is but a welcome by-product of intuitionism.

Sign: You describe your activity as mental construction, Mr. Int, but mental processes are only observable through the acts to which they lead, in your case through the words you speak and the formulas you write. Does not this mean that the only way to study intuitionism is to study the formal system which it constructs?

Int: When looking at the tree over there, I am convinced I see a tree, and it costs considerable training to replace this conviction by the knowledge that in reality lightwaves reach my eyes, leading me to the construction of an image of the tree. In the same way, in speaking to you I am convinced that I press my opinions upon you, but you instruct me that in reality I produce vibrations in the air, which cause you to perform some action, e.g. to produce other vibrations. In both cases the first view is the natural one, the second is a theoretical construction. It is too often forgotten that the truth of such constructions depends upon the present state of science and that the words "in reality" ought to be translated into "according to the contemporary view of scientists". Therefore I prefer to adhere to the idea that, when describing intuitionistic mathematics, I convey thoughts to my hearers; these words ought to be taken not in the sense of some philosophical system, but in the sense of every-day life.

Sign: Then intuitionism, as a form of interaction between men, is a social phenomenon and its study belongs to the history of civilization.

Int: Its study, not its practice. Here I agree with Mr. Prag: *primum vivere, deinde philosophari,* and if we like we can leave the latter to others. Let those who come after me wonder why I built up these mental constructions and how they can be interpreted in some philosophy; I am content to build them in the conviction that in some way they will contribute to the clarification of human thought.

Prag: It is a common fault of philosophers to speak about things they know but imperfectly and we are near to being caught in that trap. Is Mr. Int willing to give us some samples of intuitionistic reasoning, in order that we may better be able to judge the quality of the stuff?

Int: Certainly, and even I am convinced that a few lessons will give you a better insight into it than lengthy discussions. May I beg those gentlemen who are interested in my explanations, to follow me to my classroom?

Intuitionism and formalism

L. E. J. BROUWER

The subject for which I am asking your attention deals with the foundations of mathematics. To understand the development of the opposing theories existing in this field one must first gain a clear understanding of the concept "science"; for it is as a part of science that mathematics originally took its place in human thought.

By science we mean the systematic cataloguing by means of laws of nature of causal sequences of phenomena, i.e., sequences of phenomena which for individual or social purposes it is convenient to consider as repeating themselves identically, - and more particularly of such causal sequences as are of importance in social relations.

That science lends such great power to man in his action upon nature is due to the fact that the steadily improving cataloguing of ever more causal sequences of phenomena gives greater and greater possibility of bringing about desired phenomena, difficult or impossible to evoke directly, by evoking other phenomena connected with the first by causal sequences. And that man always and everywhere creates order in nature is due to the fact that he not only isolates the causal sequences of phenomena (i.e., he strives to keep them free from disturbing secondary phenomena) but also supplements them with phenomena caused by his own activity, thus making them of wider applicability. Among the latter phenomena the results of counting and measuring take so important a place, that a large number of the natural laws introduced by science treat only of the mutual relations between the results of counting and measuring. It is well to notice in this connection that a natural law in the statement of which measurable magnitudes occur can only be understood to hold in nature with a certain degree of approximation; indeed natural laws as a rule are not proof against sufficient refinement of the measuring tools.

The exceptions to this rule have from ancient times been practical arithmetic and geometry on the one hand, and the dynamics of rigid bodies and celestial mechanics on the other hand. Both these groups have so far resisted all improvements in the tools of observation. But while

Inaugural address at the University of Amsterdam, read October 14, 1912. Translated by Professor Arnold Dresden. Reprinted by the kind permission of the author and the editor from the ***Bulletin of the American Mathematical Society***, 20 (November, 1913), 81–96.

this has usually been looked upon as something accidental and temporal for the latter group, and while one has always been prepared to see these sciences descend to the rank of approximate theories, until comparatively recent times there has been absolute confidence that no experiment could ever disturb the exactness of the laws of arithmetic and geometry; this confidence is expressed in the statement that mathematics is "the" exact science.

On what grounds the conviction of the unassailable exactness of mathematical laws is based has for centuries been an object of philosophical research, and two points of view may here be distinguished, *intuitionism* (largely French) and *formalism* (largely German). In many respects these two viewpoints have become more and more definitely opposed to each other; but during recent years they have reached agreement as to this, that the exact validity of mathematical laws as laws of nature is out of the question. The question where mathematical exactness does exist, is answered differently by the two sides; the intuitionist says: in the human intellect, the formalist says: on paper.

In Kant we find an old form of intuitionism, now almost completely abandoned, in which time and space are taken to be forms of conception inherent in human reason. For Kant the axioms of arithmetic and geometry were synthetic a priori judgments, i.e., judgments independent of experience and not capable of analytical demonstration; and this explained their apodictic exactness in the world of experience as well as in abstracto. For Kant, therefore, the possibility of disproving arithmetical and geometrical laws experimentally was not only excluded by a firm belief, but it was entirely unthinkable.

Diametrically opposed to this is the view of formalism, which maintains that human reason does not have at its disposal exact images either of straight lines or of numbers larger than ten, for example, and that therefore these mathematical entities do not have existence in our conception of nature any more than in nature itself. It is true that from certain relations among mathematical entities, which we assume as axioms, we deduce other relations according to fixed laws, in the conviction that in this way we derive truths from truths by logical reasoning, but this non-mathematical conviction of truth or legitimacy has no exactness whatever and is nothing but a vague sensation of delight arising from the knowledge of the efficacy of the projection into nature of these relations and laws of reasoning. For the formalist therefore mathematical exactness consists merely in the method of developing the series of relations, and is independent of the significance one might want to give to the relations or the entities which they relate. And for the consistent formalist these meaningless series of relations to which mathematics are reduced have mathematical existence only when they have been represented in spoken or written language together with the mathematical-logical laws upon which their development depends, thus forming what is called symbolic logic.

Because the usual spoken or written languages do not in the least satisfy the requirements of consistency demanded of this symbolic logic, formalists try to avoid the use of ordinary language in mathematics. How far this may be carried is shown by the modern Italian school of formalists, whose leader, Peano, published one of his most important discoveries concerning the existence of integrals of real differential equations in the *Mathematische Annalen* in the language of symbolic logic; the result was that it could only be read by a few of the initiated and that it did not become generally available until one of these had translated the article into German.

The viewpoint of the formalist must lead to the conviction that if other symbolic formulas should be substituted for the ones that now represent the fundamental mathematical relations and the mathematical-logical laws, the absence of the sensation of delight, called "consciousness of legitimacy," which might be the result of such substitution would not in the least invalidate its mathematical exactness. To the philosopher or to the anthropologist, but not to the mathematician, belongs the task of investigating why certain systems of symbolic logic rather than others may be effectively projected upon nature. Not to the mathematician, but to the psychologist, belongs the task of explaining why we believe in certain systems of symbolic logic and not in others, in particular why we are averse to the so-called contradictory systems in which the negative as well as the positive of certain propositions are valid (Mannoury 1909: 149–54).

As long as the intuitionists adhered to the theory of Kant it seemed that the development of mathematics in the nineteenth century put them in an ever weaker position with regard to the formalists. For in the first place this development showed repeatedly how complete theories could be carried over from one domain of mathematics to another: projective geometry, for example, remained unchanged under the interchange of the rôles of point and straight line, an important part of the arithmetic of real numbers remained valid for various complex number fields and nearly all the theorems of elementary geometry remained true for non-archimedian geometry, in which there exists for every straight line segment another such segment, infinitesimal with respect to the first. These discoveries seemed to indicate indeed that of a mathematical theory only the logical form was of importance and that one need no more be concerned with the material than it is necessary to think of the significance of the digit groups with which one operates, for the correct solution of a problem in arithmetic.

But the most serious blow for the Kantian theory was the discovery of non-euclidean geometry, a consistent theory developed from a set of axioms differing from that of elementary geometry only in this respect that the parallel axiom was replaced by its negative. For this showed that the phenomena usually described in the language of elementary geometry may be described with equal exactness, though frequently less compactly in the language of non-euclidean geometry; hence, it is not only impossible to hold that the space of our experience has the properties of elementary geometry but it has no significance to ask for *the* geometry which would be true for the space of our experience. It is true that elementary geometry is better suited than any other to the description of the laws of kinematics of rigid bodies and hence of a large number of natural phenomena, but with some patience it would be possible to make objects for which the kinematics would be more easily interpretable in terms of non-euclidean than in terms of euclidean geometry (Poincaré 1903: 104).

However weak the position of intuitionism seemed to be after this period of mathematical development, it has recovered by abandoning Kant's apriority of space but adhering the more resolutely to the apriority of time. This neo-intuitionism considers the falling apart of moments of life into qualitatively different parts, to be reunited only while remaining separated by time, as the fundamental phenomenon of the human intellect, passing by abstracting from its emotional content into the fundamental phenomenon of mathematical thinking, the intuition of the bare two-oneness. This intuition of two-oneness, the basal intuition of mathematics, creates not only the numbers one and two, but also all finite ordinal numbers, inasmuch as one of the elements of the two-oneness may be thought of as a new two-oneness, which process may be repeated indefinitely; this gives rise still further to the smallest infinite ordinal number ω. Finally this basal intuition of mathematics, in which the connected and the separate, the continuous and the discrete are united, gives rise immediately to the intuition of the linear continuum, i.e., of the "between," which is not exhaustible by the interposition of new units and which therefore can never be thought of as a mere collection of units.

In this way the apriority of time does not only qualify the properties of arithmetic as synthetic a priori judgments, but it does the same for those of geometry, and not only for elementary two- and three-dimensional geometry, but for non-euclidean and n-dimensional geometries as well. For since Descartes we have learned to reduce all these geometries to arithmetic by means of the calculus of coordinates.

From the present point of view of intuitionism therefore all mathematical sets of units which are entitled to that name can be developed out of the basal intuition, and this can only be done by combining a finite number of times the two operations: "to create a finite ordinal number" and "to create the infinite ordinal number ω"; here it is to be understood that for the latter purpose any previously constructed set or any previously performed constructive operation may be taken as a unit. Consequently the intuitionist recognizes only the existence of denumerable sets, i.e., sets whose elements may be brought into one-to-one correspondence either with the elements of a finite ordinal number or with those of the infinite ordinal number ω. And in the construction of these sets neither the ordinary language nor any symbolic language can have any other rôle than that of serving as a nonmathematical auxiliary, to assist the mathematical memory or to enable different individuals to build up the same set.

For this reason the intuitionist can never feel assured of the exactness of a mathematical theory by such guarantees as the proof of its being noncontradictory, the possibility of defining its concepts by a finite number of words (Poincaré 1908a: 6), or the practical certainty that it will never lead to a misunderstanding in human relations (Borel 1912: 221).

As has been stated above, the formalist wishes to leave to the psychologist the task of selecting the "truly-mathematical" language from among the many symbolic languages that may be consistently developed. Inasmuch as psychology has not yet begun in this task, formalism is compelled to mark off, at least temporarily, the domain that it wishes to consider as "true mathematics" and to lay down for that purpose a definite system of axioms and laws of reasoning, if it does not wish to see its work doomed to sterility. The various ways in which this attempt has actually been made all follow the same leading idea, viz., the presupposition of the existence of a world of mathematical objects, a world independent of the thinking individual, obeying the laws of classical logic and whose objects may possess with respect to each other the "relation of a set to its elements." With reference to this relation various axioms are postulated, suggested by the practice with natural finite sets; the principal of these are: *"a set is determined by its elements"*; *"for any two mathematical objects it is decided whether or not one of them is contained in the other one as an element"*; *"to every set belongs another set containing as its elements nothing but the subsets of the given set"*; the axiom of selection: *"a set which is split into subsets contains at least one subset which contains one and not more than one element of each of the first subsets"*; the axiom of inclusion: *"if for any mathematical object it is decided whether a certain property is valid for it or not, then there exists a set containing* *nothing but those objects for which the property does hold"*; the axiom of composition: *"the elements of all sets that belong to a set of sets form a new set."*

On the basis of such a set of axioms the formalist develops now in the first place the theory of "finite sets." A set is called finite if its elements cannot be brought into one-to-one correspondence with the elements of one of its subsets; by means of relatively complicated reasoning the principle of complete induction is proved to be a fundamental property of these sets (Zermelo 1909: 185–93); this principle states that a property will be true for all finite sets if, first, it is true for all sets containing a single element, and, second, its validity for an arbitrary finite set follows from its validity for this same set reduced by a single one of its elements. That the formalist must give an explicit proof of this principle, which is self-evident for the finite numbers of the intuitionist on account of their construction, shows at the same time that the former will never be able to justify his choice of axioms by replacing the unsatisfactory appeal to inexact practice or to intuition equally inexact for him by a proof of the non-contradictoriness of his theory. For in order to prove that a contradiction can never arise among the infinitude of conclusions that can be drawn from the axioms he is using, he would first have to show that if no contradiction had as yet arisen with the nth conclusion then none could arise with the $(n+1)$th conclusion, and secondly, he would have to apply the principle of complete induction intuitively. But it is this last step which the formalist may not take, even though he should have proved the principle of complete induction; for this would require mathematical certainty that the set of properties obtained after the nth conclusion had been reached, would satisfy for an arbitrary n his definition for finite sets (Poincaré 1905: 834), and in order to secure this certainty he would have to have recourse not only to the unpermissible application of a symbolic criterion to a concrete example but also to another intuitive application of the principle of complete induction; this would lead him to a vicious circle reasoning.

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. In the course of these

formalistic developments it turns out that the consistent application of the axiom of inclusion leads inevitably to contradictions. A clear illustration of this fact is furnished by the so-called paradox of Burali-Forti (1897). To exhibit it we have to lay down a few definitions.

A set is called ordered if there exists between any two of its elements a relation of "higher than" or "lower than," with this understanding that if the element a is higher than the element b, then the element b is lower than the element a, and if the element b is higher than a and c is higher than b, then c is higher than a.

A well-ordered set (in the formalistic sense) is an ordered set, such that every subset contains an element lower than all others.

Two well-ordered sets that may be brought into one-to-one correspondence under invariance of the relations of "higher than" and "lower than" are said to have the same ordinal number.

If two ordinal numbers A and B are not equal, then one of them is greater than the other one, let us say A is greater than B; this means that B may be brought into one-to-one correspondence with an initial segment of A under invariance of the relations of "higher than" and "lower than." We have introduced above, from the intuitionist viewpoint, the smallest infinite ordinal number ω, i.e., the ordinal number of the set of all finite ordinal numbers arranged in order of magnitude.[1] Well-ordered sets having the ordinal number ω are called elementary series.

It is proved without difficulty by the formalist that an arbitrary subset of a well-ordered set is also a well-ordered set, whose ordinal number is less than or equal to that of the original set; also, that if to a well-ordered set that does not contain all mathematical objects a new element be added that is defined to be higher than all elements of the original set, a new well-ordered set arises whose ordinal number is greater than that of the first set.

We construct now on the basis of the axiom of inclusion the *set s which contains as elements all the ordinal numbers arranged in order of magnitude;* then we can prove without difficulty, on the one hand that s is a well-ordered set whose ordinal number can not be exceeded by any other ordinal number in magnitude, and on the other hand that it is possible, since not all mathematical objects are ordinal numbers, to create an ordinal number greater than that of s by adding a new element to s, – a contradiction.[2]

[1]The more general ordinal numbers of the intuitionist are the numbers constructed by means of Cantor's two principles of generation (cf. Cantor 1895–7, 49: 226).

[2]It is without justice that the paradox of Burali-Forti is sometimes classed with that of Richard, which in a somewhat simplified form reads as follows: "Does there exist a *least integer, that can not be defined by a sentence of at most twenty words?* On the one hand *yes*, for the number of sentences of at most twenty words is of course finite; on the other

Although the formalists must admit contradictory results as mathematical if they want to be consistent, there is something disagreeable for them in a paradox like that of Burali-Forti because at the same time the progress of their arguments is guided by the principium contradictionis, i.e., by the rejection of the simultaneous validity of two contradictory properties. For this reason the axiom of inclusion has been modified to read as follows: *"If for all elements of a set it is decided whether a certain property is valid for them or not, then the set contains a subset containing nothing but those elements for which the property does hold"* (Zermelo 1908: 263).

In this form the axiom permits only the introduction of such sets as are subsets of sets previously introduced; if one wishes to operate with other sets, their existence must be explicitly postulated. Since however in order to accomplish anything at all the existence of a certain collection of sets will have to be postulated at the outset, the only valid argument that can be brought against the introduction of a new set is that it leads to contradictions; indeed the only modifications that the discovery of paradoxes has brought about in the practice of formalism has been the abolition of those sets that had given rise to these paradoxes. One continues to operate without hesitation with other sets introduced on the basis of the old axiom of inclusion; the result of this is that extended fields of research, which are without significance for the intuitionist are still of considerable interest to the formalist. An example of this is found in the theory of potencies, of which I shall sketch the principal features here, because it illustrates so clearly the impassable chasm which separates the two sides.

Two sets are said to possess the same potency, or power, if their elements can be brought into one-to-one correspondence. The power of set A is said to be greater than that of B, and the power of B less than that of A, if it is possible to establish a one-to-one correspondence between B and a part of A, but impossible to establish such a correspondence between A and a part of B. The power of a set which has the same power as one of its subsets, is called infinite, other powers are called finite. Sets that have the same power as the oridinal number ω are called denumerably infinite and the power of such sets is called aleph-null: it proves to be the smallest infinite power. According to the statements previously made, this power aleph-null is the only infinite power of which the intuitionists recognize the existence.

hand *no,* for if it should exist, it would be defined by the sentence of fifteen words formed by the words italicized above."

The origin of this paradox does not lie in the axiom of inclusion but in the variable meaning of the word *"defined"* in the italicized sentence, which makes it possible to define by means of this sentence an infinite number of integers in succession.

Let us now consider the concept: "denumerably infinite ordinal number." From the fact that this concept has a clear and well-defined meaning for both formalist and intuitionist, the former infers the right to create the "set of all denumerably infinite ordinal numbers," the power of which he calls aleph-one, a right not recognized by the intuitionist. Because it is possible to argue to the satisfaction of both formalist and intuitionist, first, that denumerably infinite sets of denumerably infinite ordinal numbers can be built up in various ways, and second, that for every such set it is possible to assign a denumerably infinite ordinal number, ***not*** belonging to this set, the formalist concludes: "aleph-one is greater than aleph-null," a proposition that has no meaning for the intuitionist. Because it is possible to argue to the satisfaction of both formalist and intuitionist that it is impossible to construct[3] a set of denumerably infinite ordinal numbers, which could be proved to have a power less than that of aleph-one, but greater than that of aleph-null, the formalist concludes: "aleph-one is the second smallest infinite ordinal number," a proposition that has no meaning for the intuitionist.

Let us consider the concept: "real number between 0 and 1." For the formalist this concept is equivalent to "elementary series of digits after the decimal point,"[4] for the intuitionist it means "law for the construction of an elementary series of digits after the decimal point, built up by means of a finite number of operations." And when the formalist creates the "set of all real numbers between 0 and 1," these words are without meaning for the intuitionist, even whether one thinks of the real numbers of the formalist, determined by elementary series of freely selected digits, or of the real numbers of the intuitionist, determined by finite laws of construction. Because it is possible to prove to the satisfaction of both formalist and intuitionist, first, that denumerably infinite sets of real numbers between 0 and 1 can be constructed in various ways, and second that for every such set it is possible to assign a real number between 0 and 1, not belonging to the set, the formalist concludes: "the power of the continuum, i.e., the power of the set of real numbers between 0 and 1, is greater than aleph-null," a proposition that is without meaning for the intuitionist; the formalist further raises the question, whether there exist sets of real numbers between 0 and 1, whose power is less than that of the continuum, but greater than aleph-null, in other words, "whether the power of the continuum is the second smallest infinite power," and this

[3]If "construct" were here replaced by "define" (in the formalistic sense), the proof would *not* be satisfactory to the intuitionist. For, in Cantor's argument it is not allowed to replace the words "können wir bestimmen" (1895–7, 49: 214, line 17 from top) by the words "muss es geben."

[4]Here as everywhere else in this paper, the assumption is tacitly made that there are an infinite number of digits different from 9.

question, which is still waiting for an answer, he considers to be one of the most difficult and most fundamental of mathematical problems.

For the intuitionist, however, the question as stated is without meaning; and as soon as it has been so interpreted as to get a meaning, it can easily be answered.

If we restate the question in this form: "Is it impossible to construct[5] infinite sets of real numbers between 0 and 1, whose power is less than that of the continuum, but greater than aleph-null?" then the answer must be in the affirmative; for the intuitionist can only construct denumerable sets of mathematical objects and if, on the basis of the intuition of the linear continuum, he admits elementary series of free selections as elements of construction, then each non-denumerable set constructed by means of it contains a subset of the power of the continuum.

If we restate the question in the form: "Is it possible to establish a one-to-one correspondence between the elements of a set of denumerably infinite ordinal numbers on the one hand, and a set of real numbers between 0 and 1 on the other hand, both sets being indefinitely extended by the construction of new elements, of such a character that the correspondence shall not be disturbed by any continuation of the construction of both sets?" then the answer must also be in the affirmative, for the extension of both sets can be divided into phases in such a way as to add a denumerably infinite number of elements during each phase.[6]

If however we put the question in the following form: "Is it possible to construct a law which will assign a denumerably infinite ordinal number to every elementary series of digits and which will give certainty a priori that two different elementary series will never have the same denumerably infinite ordinal number corresponding to them?" then the answer must be in the negative; for this law of correspondence must prescribe in some way a construction of certain denumerably infinite ordinal numbers at each of the successive places of the elementary series; hence there is for each place c_ν a well-defined largest denumerably infinite number α_ν, the construction of which is suggested by that particular place; there is then also a well-defined denumerably infinite ordinal number α_ω,

[5]If "construct" were here replaced by "define" (in the formalistic sense), and if we suppose that the problem concerning the pairs of digits in the decimal fraction development of π, discussed on p. 88, *can not be solved,* then the question of the text must be answered negatively. For, let us denote by Z the set of those infinite binary fractions, whose nth digit is 1, if the nth pair of digits in the decimal fraction development of π consists of unequal digits; let us further denote by X the set of all finite binary fractions. Then the power of $Z+X$ is greater than aleph-null, but less than that of the continuum.

[6]Calling *denumerably unfinished* all sets of which the elements can be individually realized, and in which for every denumerably infinite subset there exists an element not belonging to this subset, we can say in general, in accordance with the definitions of the text: *"All denumerably unfinished sets have the same power."*

greater than all α_ν's and that can not therefore be exceeded by any of the ordinal numbers involved by the law of correspondence; hence the power of that set of ordinal numbers can not exceed aleph-null.

As a means for obtaining ever greater powers, the formalists define with every power μ a "set of all the different ways in which a number of selections of power μ may be made," and they prove that the power of this set is greater than μ. In particular, when it has been proved to the satisfaction of both formalist and intuitionist that it is possible in various ways to construct laws according to which functions of a real variable different from each other are made to correspond to all elementary series of digits, but that it is impossible to construct a law according to which an elementary series of digits is made to correspond to every function of a real variable and in which there is certainty a priori that two different functions will never have the same elementary series corresponding to them, the formalist concludes: "the power c' of the set of all functions of a real variable is greater than the power c of the continuum," a proposition without meaning to the intuitionist; and in the same way in which he was led from c to c', he comes from c' to a still greater power c''.

A second method used by the formalists for obtaining ever greater powers is to define for every power μ, which can serve as a power of ordinal numbers, "the set of all ordinal numbers of power μ," and then to prove that the power of this set is greater than μ. In particular they denote by aleph-two the power of the set of all ordinal numbers of power aleph-one and they prove that aleph-two is greater than aleph-one and that it follows in magnitude immediately after aleph-one. If it should be possible to interpret this result in a way in which it would have meaning for the intuitionist, such interpretation would not be as simple in this case as it was in the preceding cases.

What has been treated so far must be considered to be the negative part of the theory of potencies; for the formalist there also exists a positive part however, founded on the theorem of Bernstein: "If the set A has the same power as a subset of B and B has the same power as a subset of A, then A and B have the same power" or, in an equivalent form: "If the set $A = A_1 + B_1 + C_1$, has the same power as the set A_1, then it also has the same power as the set $A_1 + B_1$.

This theorem is self-evident for denumerable sets. If it is to have any meaning at all for sets of higher power for the intuitionist, it will have to be interpretable as follows: "If it is possible, *first* to construct a law determining a one-to-one correspondence between the mathematical entities of type A and those of type A_1, and *second* to construct a law determining a one-to-one correspondence between the mathematical entities of type A and those of types A_1, B_1, and C_1, then it is possible to determine from these two laws by means of a finite number of operations a third law, determining a one-to-one correspondence between the mathematical entities of type A and those of types A_1 and B_1."

In order to investigate the validity of this interpretation, we quote the proof:

"From the division of A into $A_1+B_1+C_1$, we secure by means of the correspondence γ_1 between A and A_1 a division of A_1 into $A_2+B_2+C_2$, as well as a one-to-one correspondence γ_2 between A_1 and A_2. From the division of A_1 into $A_2+B_2+C_2$, we secure by means of the correspondence between A_1 and A_2 a division of A_2 into $A_3+B_3+C_3$, as well as a one-to-one correspondence γ_3 between A_2 and A_3. Indefinite repetition of this procedure will divide the set A into an elementary series of subsets $C_1, C_2, C_3, \ldots$, an elementary series of subsets $B_1, B_2, B_3, \ldots$, and a remainder set D. The correspondence γ_C between A and A_1+B_1 which is desired is secured by assigning to every element of C_ν the corresponding element of $C_{\nu+1}$ and by assigning every other element of A to itself."

In order to test this proof on a definite example, let us take for A the set of all real numbers between 0 and 1, represented by infinite decimal fractions, for A_1 the set of those decimal fractions in which the $(2n-1)$th digit is equal to the $2n$th digit; further a decimal fraction that does not belong to A_1 will be counted to belong to B_1 or to C_1 according as the above-mentioned equality of digits occurs an infinite or a finite number of times. By replacing successively each digit of an arbitrary element of A by a pair of digits equal to it, we secure at once a law determining a one-to-one correspondence γ_1 between A and A_1. For of the element of A_1 that corresponds to an arbitrary well-defined element of A, such as, e.g., $\pi-3$, we can determine successively as many digits as we please; it must therefore be considered as being well-defined.

In order to determine the element corresponding to $\pi-3$ according to the correspondence γ_C, it is now necessary to decide first whether it happens an infinite or a finite number of times in the decimal fraction development of $\pi-3$ that a digit in an odd-numbered place is equal to the digit in the following even-numbered place; for this purpose we should either have to invent a process for constructing an elementary series of such pairs of equal digits, or to deduce a contradiction from the assumption of the existence of such an elementary series. There is, however, no ground for believing that either of these problems can be solved.[7]

[7]Such belief could be based only on an appeal to the principium tertii exclusi, i.e., to the axiom of the existence of the "set of all mathematical properties," an axiom of far wider range even than the axioms of inclusion, quoted above. Compare in this connection Brouwer 1908: 152–8.

Hence it has become evident that also the theorem of Bernstein, and with it the positive part of the theory of potencies, does not allow an intuitionistic interpretation.

So far my exposition of the fundamental issue, which divides the mathematical world. There are eminent scholars on both sides and the chance of reaching an agreement within a finite period is practically excluded. To speak with Poincaré: "Les hommes ne s'entendent pas, parce qu'ils ne parlent pas la même langue et qu'il y a des langues qui ne s'apprennent pas."

Consciousness, philosophy, and mathematics

L. E. J. BROUWER

The ... point of view that there are no non-experienced truths and that logic is not an absolutely reliable instrument to discover truths has found acceptance with regard to mathematics much later than with regard to practical life and to science. Mathematics rigorously treated from this point of view, including deducing theorems exclusively by means of introspective construction, is called intuitionistic mathematics. In many respects it deviates from classical mathematics. In the first place because classical mathematics uses logic to generate theorems, believes in the existence of unknown truths, and in particular applies the *principle of the excluded third* expressing that every mathematical assertion (i.e. every assignment of a mathematical property to a mathematical entity) either is a truth or cannot be a truth. In the second place because classical mathematics confines itself to *predeterminate* infinite sequences for which from the beginning the nth element is fixed for each n. Owing to this confinement classical mathematics, to define real numbers, has only predeterminate convergent infinite sequences of rational numbers at its disposal. Out of real numbers defined in this way, only subspecies of "ever unfinished denumerable" species of real numbers can be composed by means of introspective construction. Such ever unfinished denumerable species all being of measure zero, classical mathematics, to create the continuum out of points, needs some logical process starting from one or more axioms. Consequently we may say that classical analysis, however appropriate it be for technique and science, has less mathematical truth than intuitionistic analysis performing the said composition of the continuum by considering the species of freely proceeding convergent infinite sequences of rational numbers, without having recourse to language or logic.

As a matter of course also the languages of the two mathematical schools diverge. And even in those mathematical theories which are covered by a neutral language, i.e. by a language understandable on both sides, either school operates with mathematical entities not recognized

Excerpted by kind permission of the publisher from 10th International Congress of Philosophy, Amsterdam, 1948, *Proceedings I*, Fascicule II (Amsterdam: North-Holland Publishing Company, 1949), pp. 1243–9.

by the other one: there are intuitionist structures which cannot be fitted into any classical logical frame, and there are classical arguments not applying to any introspective image. Likewise, in the theories mentioned, mathematical entities recognized by both parties on each side are found satisfying theorems which for the other school are either false, or senseless, or even in a way contradictory. In particular, theorems holding in intuitionism, but not in classical mathematics, often originate from the circumstance that for mathematical entities belonging to a certain species, the possession of a certain property imposes a special character on their way of development from the basic intuition, and that from this special character of their way of development from the basic intuition, properties ensue which for classical mathematics are false. A striking example is the intuitionist theorem that a full function of the unity continuum, i.e. a function assigning a real number to every non-negative real number not exceeding unity, is necessarily uniformly continuous.

To elucidate the consequences of the rejection of the principle of the excluded third as an instrument to discover truths, we shall put the wording of this principle into the following slightly modified, intuitionistically more adequate form, called the *simple principle of the excluded third:*

> *Every assignment τ of a property to a mathematical entity can be judged, i.e. either proved or reduced to absurdity.*

Then for a single such assertion τ the enunciation of this principle is non-contradictory in intuitionistic as well as in classical mathematics. For, if it were contradictory, then the absurdity of τ would be true and absurd at the same time, which is impossible. Moreover, as can easily be proved, for a *finite* number of such assertions τ the simultaneous enunciation of the principle is non-contradictory likewise. However, for the simultaneous enunciation of the principle for all elements of an *arbitrary* species of such assertions τ this non-contradictority cannot be maintained.

E.g. from the supposition, for a definite real number c_1, that the assertion: c_1 *is rational,* has been proved to be either true or contradictory, no contradiction can be deduced. Furthermore, $c_1, c_2, \ldots c_m$ being real numbers, neither the simultaneous supposition, for each of the values $1, 2, \ldots m$ of ν, that the assertion: c_ν *is rational,* has been proved to be either true or contradictory, can lead to a contradiction. However, the simultaneous supposition for *all* real numbers c that the assertion: c *is rational,* has been proved to be either true or contradictory, does lead to a contradiction.

Consequently if we formulate the *complete principle of the excluded third* as follows:

If a, b, and c are species of mathematical entities, if further both a and b form part of c, and if b consists of those elements of c which cannot belong to a, then c is identical with the union of a and b,

the latter principle is contradictory.

A corollary of the *simple* principle of the excluded third says that:

If for an assignment τ of a property to a mathematical entity the non-contradictority, i.e. the absurdity of the absurdity, has been established, the truth of τ can be demonstrated likewise.

The analogous corollary of the *complete* principle of the excluded third is the *principle of reciprocity of complementarity,* running as follows:

If a, b, and c are species of mathematical entities, if further a and b form part of c, and if b consists of the elements of c which cannot belong to a, then a consists of the elements of c which cannot belong to b.

Another corollary of the *simple* principle of the excluded third is the *simple principle of testability* saying that

every assignment τ of a property to a mathematical entity can be tested, i.e. proved to be either non-contradictory or absurd.

The analogous corollary of the *complete* principle of the excluded third is the following *complete principle of testability:*

If a, b, d, and c are species of mathematical entities, if each of the species a, b, and d forms part of c, if b consists of the elements of c which cannot belong to a, and d of the elements of c which cannot belong to b, then c is identical with the union of b and d.

For intuitionism the principle of the excluded third and its corollaries are assertions σ about assertions τ, and these assertions σ only then are "realized", i.e. only then convey truths, if these truths have been experienced.

Each assertion τ of the possibility of a construction of bounded finite character in a finite mathematical system furnishes a case of realization of the principle of the excluded third. For every such construction can be attempted only in a finite number of particular ways, and each attempt proves successful or abortive in a finite number of steps.

If the assertion of an absurdity is called a *negative assertion,* then each negative assertion furnishes a case of realization of the principle of reciprocity of complementarity. For, let α be a negative assertion, indicating

the absurdity of the assertion β. As, on the one hand, the implication of the truth of an assertion a by the truth of an assertion b implies the implication of the absurdity of b by the absurdity of a, whilst, on the other hand, the truth of β implies the absurdity of the absurdity of β, we conclude that the absurdity of the absurdity of the absurdity of β, i.e. the non-contradictority of α, implies the absurdity of β, i.e. implies α.

In consequence of this realization of the principle of reciprocity of complementarity the principles of testability and of the excluded third are equivalent in the domain of negative assertions. For, if for α the principle of testability holds, this means that either the absurdity of the absurdity of β or the non-contradictority of the absurdity of β, i.e. by the preceding paragraph, that either the absurdity of the absurdity of β or the absurdity of β, i.e. either the absurdity of α or α can be proved, so that α satisfies the principle of the excluded third.

To give some examples refuting the principle of the excluded third and its corollaries, we introduce the notion of a *drift.* By a drift we understand the union γ of a convergent fundamental sequence of real numbers $c_1(\gamma), c_2(\gamma), \ldots$, called the *counting-numbers* of the drift, and the limiting-number $c(\gamma)$ of this sequence, called the *kernel* of the drift, all counting-numbers lying apart[1] from each other and from the kernel. If $c_\nu(\gamma) <\circ c(\gamma)$ for each ν, the drift will be called *left-winged.* If $c_\nu(\gamma) \circ> c(\gamma)$ for each ν, the drift will be called *right-winged.* If the fundamental sequence $c_1(\gamma), c_2(\gamma), \ldots$ is the union of a fundamental sequence of *left counting-numbers* $l_1(\gamma), l_2(\gamma), \ldots$ such that $l_\nu(\gamma) <\circ c(\gamma)$ for each ν, and a fundamental sequence of *right counting-numbers* $d_1(\gamma), d_2(\gamma), \ldots$ such that $d_\nu(\gamma) \circ> c(\gamma)$ for each ν, the drift will be called *two-winged.*

Let α be a mathematical assertion so far neither tested nor recognized as testable. Then in connection with this assertion α and with a drift γ the creating subject can generate an infinitely proceeding sequence $R(\gamma, \alpha)$ of real numbers $c_1(\gamma, \alpha), c_2(\gamma, \alpha), \ldots$ according to the following direction: As long as during the choice of the $c_n(\gamma, \alpha)$ the creating subject has experienced neither the truth, nor the absurdity of α, each $c_n(\gamma, \alpha)$ is chosen equal to $c(\gamma)$. But as soon as between the choice of $c_{r-1}(\gamma, \alpha)$ and that of $c_r(\gamma, \alpha)$ the creating subject has experienced either the truth or the absurdity of α, $c_r(\gamma, \alpha)$, and likewise $c_{r+\nu}(\gamma, \alpha)$ for each natural

[1]If for two real numbers a and b defined by convergent infinite sequences of rational numbers $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ respectively, two such natural numbers m and n can be calculated that $b_\nu - a_\nu > 2^{-n}$ for $\nu \geqslant m$, we write $b \circ> a$ and $a <\circ b$, and a and b are said to lie *apart* from each other. If $a = b$ is absurd, we write $a \neq b$. If $a <\circ b$ is absurd, we write $a \ngtr b$. If both $a = b$ and $a <\circ b$ are absurd, we write $a > b$. The absurdities of $a <\circ b$ and $a < b$ prove to be mutually equivalent, and the absurdity of $a \ngtr b$ proves to be equivalent to $a < b$.

number ν, is chosen equal to $c_r(\gamma)$. This sequence $R(\gamma, \alpha)$ converges to a real number $D(\gamma, \alpha)$ which will be called a ***direct checking-number of*** γ ***through*** α.

Again, in connection with α and with a two-winged drift γ the creating subject can generate an infinitely proceeding sequence $S(\gamma, \alpha)$ of real numbers $\omega_1(\gamma, \alpha), \omega_2(\gamma, \alpha), \ldots$ according to the following direction: As long as during the choice of the $\omega_n(\gamma, \alpha)$ the creating subject has experienced neither the truth, nor the absurdity of α, each $\omega_n(\gamma, \alpha)$ is chosen equal to $c(\gamma)$. But as soon as between the choice of $\omega_{r-1}(\gamma, \alpha)$ and that of $\omega_r(\gamma, \alpha)$ the creating subject has experienced the truth of α, $\omega_r(\gamma, \alpha)$, and likewise $\omega_{r+\nu}(\gamma, \alpha)$ for each natural number ν, is chosen equal to $d_r(\gamma)$. And as soon as between the choice of $\omega_{s-1}(\gamma, \alpha)$ and that of $\omega_s(\gamma, \alpha)$ the creating subject has experienced the absurdity of α, $\omega_s(\gamma, \alpha)$, and likewise $\omega_{s+\nu}(\gamma, \alpha)$ for each natural number ν, is chosen equal to $l_s(\gamma)$. This sequence $S(\gamma, \alpha)$ converges to a real number $E(\gamma, \alpha)$ which will be called an ***oscillatory checking-number of*** γ ***through*** α.

Let γ be a right-winged drift whose counting-numbers are rational. Then the assertion of the rationality of $D(\gamma, \alpha)$ is testable, but not judgable, and its non-contradictority is not equivalent to its truth. Furthermore we have $D(\gamma, \alpha) > c(\gamma)$, but not $D(\gamma, \alpha) \circ\!> c(\gamma)$.

Let γ be a two-winged drift whose right counting-numbers are rational, and whose left counting-numbers are irrational. Then the assertion of the rationality of $E(\gamma, \alpha)$ is neither judgeable, nor is it testable, nor is its non-contradictority equivalent to its truth. Furthermore $E(\gamma, \alpha)$ is neither $\geqslant c(\gamma)$, nor $\leqslant c(\gamma)$.

The long belief in the universal validity of the principle of the excluded third in mathematics is considered by intuitionism as a phenomenon of history of civilization of the same kind as the old-time belief in the rationality of π or in the rotation of the firmament on an axis passing through the earth. And intuitionism tries to explain the long persistence of this dogma by two facts: firstly the obvious non-contradictority of the principle for an arbitrary single assertion; secondly the practical validity of the whole of classical logic for an extensive group of *simple everyday phenomena*. The latter fact apparently made such a strong impression that the play of thought that classical logic originally was, became a deep-rooted habit of thought which was considered not only as useful but even as aprioristic.

Obviously the field of validity of the principle of the excluded third is identical with the intersection of the fields of validity of the principle of testability and the principle of reciprocity of complementarity. Furthermore the former field of validity is a *proper* subfield of each of the latter ones, as is shown by the following examples:

Let A be the species of the direct checking-numbers of drifts with rational counting-numbers, B the species of the irrational real numbers, C the union of A and B. Then all assertions of rationality of an element of C satisfy the principle of testability, whilst there are assertions of rationality of an element of C not satisfying the principle of the excluded third. Again, all assertions of equality of two real numbers satisfy the principle of reciprocity of complementarity, whereas there are assertions of equality of two real numbers not satisfying the principle of the excluded third.

In the domain of mathematical assertions the property of absurdity, just as the property of truth, is a *universally additive property*, that is to say, if it holds for each element α of a species of assertions, it also holds for the assertion which is the union of the assertions α. *This property of universal additivity does not obtain for the property of non-contradictority.* However, non-contradictority does possess the weaker property of *finite additivity*, that is to say, if the assertions ρ and σ are non-contradictory, the assertion τ which is the union of ρ and σ, is also non-contradictory. For, let us start for a moment from the supposition ω that τ is contradictory. Then the truth of ρ would entail the contradictority of σ, which would clash with the data, so that the truth of ρ is absurd, i.e. ρ is absurd. This consequence of the supposition ω clashing with the data, the supposition ω is contradictory, i.e. τ is non-contradictory.

Application of this theorem to the special non-contradictory assertions that are the enunciations of the principle of the excluded third for a single assertion, establishes the above-mentioned non-contradictority of the simultaneous enunciation of this principle for a finite number of assertions.

Within some species of mathematical entities the absurdities of two non-equivalent[2] assertions may be equivalent. E.g. each of the following three pairs of non-equivalent assertions relative to a real number a:

I 1. $a=a$;	I 2. either $a \leqslant 0$ or $a \geqslant 0$
II 1. $a \geqslant 0$;	II 2. either $a=0$ or $a \circ\!> 0$
III 1. $a>0$;	III 2. $a \circ\!> 0$

furnishes a pair of equivalent absurdities.

It occurs that within some species of mathematical entities some absurdities of constructive properties can be given a constructive form. E.g. for a natural number a the absurdity of the existence of two natural numbers different from a and from 1 and having a as their product is equivalent to the existence, whenever a is divided by a natural number dif-

[2]By non-equivalence we understand absurdity of equivalence, just as by noncontradictority we understand absurdity of contradictority.

ferent from a and from 1, of a remainder. Likewise, for two real numbers a and b the relation $a \geqslant b$ introduced above as an absurdity of a constructive property can be formulated constructively as follows: Let $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ be convergent infinite sequences of rational numbers defining a and b respectively. Then, for any natural number n, a natural number m can be calculated such that $a_\nu - b_\nu \circ\!> -2^{-n}$ for $\nu \geqslant m$.

On the other hand there seems to be little hope for reducing irrationality of a real number a, or one of the relations $a \neq b$ and $a > b$ for real numbers a and b, to a constructive property, if we remark that a direct checking-number of a drift whose kernel is rational and whose counting-numbers are irrational, is irrational without lying apart from the species of rational numbers; further that a direct checking-number of an arbitrary drift differs from the kernel of the drift without lying apart from it, and that a direct checking-number of a right-winged drift lies to the right of the kernel of the drift without lying apart from it.

It occurs that within some species of mathematical entities some non-contradictorities of constructive properties ζ can be given either a constructive form (possibly, but not necessarily, in consequence of reciprocity of complementarity holding for ζ) or the form of an absurdity of a constructive property. E.g. for real numbers a and b the non-contradictority of $a = b$ is equivalent to $a = b$, and the non-contradictority of: *either* $a = b$ *or* $a \circ\!> b$, is equivalent to $a \geqslant b$; further the non-contradictority of $a \circ\!> b$ is equivalent to the absurdity of $a \leqslant b$ as well as to the absurdity of: *either* $a = b$ or $a <\!\circ\, b$.

On the other hand, if we think of the property of non-contradictority of rationality existing for all direct checking-numbers of drifts whose counting-numbers are rational, there seems to be little hope for reducing non-contradictority of rationality of a real number to a constructive property or to an absurdity of a constructive property.

If we understand by the *simple absurdity* of the property η the absurdity of η, and by the *$(n+1)$-fold absurdity* of η the absurdity of the n-fold absurdity of η, then a theorem established above expresses that *threefold absurdity is equivalent to simple absurdity*. And a corollary of this theorem is that *n-fold absurdity is equivalent to simple or to double absurdity according as n is odd or even.*

I should like to terminate here. I hope I have made clear that intuitionism on the one hand subtilizes logic, on the other hand denounces logic as a source of truth. Further that intuitionistic mathematics is inner architecture, and that research in foundations of mathematics is inner inquiry with revealing and liberating consequences, also in non-mathematical domains of thought.

The philosophical basis of intuitionistic logic

MICHAEL DUMMETT

The question with which I am here concerned is: What plausible rationale can there be for repudiating, within mathematical reasoning, the canons of classical logic in favour of those of intuitionistic logic? I am, thus, not concerned with justifications of intuitionistic mathematics from an eclectic point of view, that is, from one which would admit intuitionistic mathematics as a legitimate and interesting form of mathematics alongside classical mathematics: I am concerned only with the standpoint of the intuitionists themselves, namely that classical mathematics employs forms of reasoning which are not valid on any legitimate way of construing mathematical statements (save, occasionally, by accident, as it were, under a quite unintended reinterpretation). Nor am I concerned with exegesis of the writings of Brouwer or of Heyting: the question is what forms of justification of intuitionistic mathematics will stand up, not what particular writers, however eminent, had in mind. And, finally, I am concerned only with the most fundamental feature of intuitionistic mathematics, its underlying logic, and not with the other respects (such as the theory of free choice sequences) in which it differs from classical mathematics. It will therefore be possible to conduct the discussion wholly at the level of elementary number theory. Since we are, in effect, solely concerned with the logical constants – with the sentential operators and the first-order quantifiers – our interest lies only with the most general features of the notion of a mathematical construction, although it will be seen that we need to consider these in a somewhat delicate way.

Any justification for adopting one logic rather than another as the logic for mathematics must turn on questions of *meaning*. It would be impossible to contrive such a justification which took meaning for granted, and represented the question as turning on knowledge or certainty. We are certain of the truth of a statement when we have conclusive grounds for it and are certain that the grounds which we have *are* valid grounds for it and *are* conclusive. If classical arguments for mathematical statements are called in question, this cannot possibly be because

Reprinted with the kind permission of the author, the editors, and the publisher from *Proceedings of the Logic Colloquium, Bristol, July 1973*, H. E. Rose and J. C. Shepherdson, eds., North-Holland 1975, pp. 5–40.

it is thought that we are, in general, unable to tell with certainty whether an argument is classically valid, unless it is also intuitionistically valid: rather, it must be that what is being put in doubt is whether arguments which are valid by classical but not by intuitionistic criteria are absolutely valid, that is, whether they really do conclusively establish their conclusions as true. Even if it were held that classical arguments, while not in general absolutely valid, nevertheless always conferred a high probability on their conclusions, it would be wrong to characterise the motive for employing only intuitionistic arguments as lying in a desire to attain knowledge in place of mere probable opinion in mathematics, since the very thesis that the use of classical arguments did not lead to knowledge would represent the crucial departure from the classical conception, beside which the question of whether or not one continued to make use of classical arguments as mere probabilistic reasoning is comparatively insignificant. (In any case, within standard intuitionistic mathematics, there is no reason whatever why the existence of a classical proof of it should render a statement probable, since if, e.g., it is a statement of analysis, its being a classical theorem does not prevent it from being intuitionistically disprovable.)

So far as I am able to see, there are just two lines of argument for repudiating classical reasoning in mathematics in favour of intuitionistic reasoning. The first runs along the following lines. The meaning of a mathematical statement determines and is exhaustively determined by its *use*. The meaning of such a statement cannot be, or contain as an ingredient, anything which is not manifest in the use made of it, lying solely in the mind of the individual who apprehends that meaning: if two individuals agree completely about the use to be made of the statement, then they agree about its meaning. The reason is that the meaning of a statement consists solely in its rôle as an instrument of communication between individuals, just as the powers of a chess-piece consist solely in its rôle in the game according to the rules. An individual cannot communicate what he cannot be observed to communicate: if one individual associated with a mathematical symbol or formula some mental content, where the association did not lie in the use he made of the symbol or formula, then he could not convey that content by means of the symbol or formula, for his audience would be unaware of the association and would have no means of becoming aware of it.

The argument may be expressed in terms of the *knowledge* of meaning, i.e. of understanding. A model of meaning is a model of understanding, i.e. a representation of what it is that is known when an individual knows the meaning. Now knowledge of the meaning of a particular symbol or expression is frequently verbalisable knowledge, that is, knowledge which

consists in the ability to state the rules in accordance with which the expression or symbol is used or the way in which it may be replaced by an equivalent expression or sequence of symbols. But to suppose that, in general, a knowledge of meaning consisted in verbalisable knowledge would involve an infinite regress: if a grasp of the meaning of an expression consisted, in general, in the ability to *state* its meaning, then it would be impossible for anyone to learn a language who was not already equipped with a fairly extensive language. Hence that knowledge which, in general, constitutes the understanding of the language of mathematics must be implicit knowledge. Implicit knowledge cannot, however, meaningfully be ascribed to someone unless it is possible to say in what the manifestation of that knowledge consists: there must be an observable difference between the behaviour or capacities of someone who is said to have that knowledge and someone who is said to lack it. Hence it follows, once more, that a grasp of the meaning of a mathematical statement must, in general, consist of a capacity to use that statement in a certain way, or to respond in a certain way to its use by others.

Another approach is via the idea of learning mathematics. When we learn a mathematical notation, or mathematical expressions, or, more generally, the language of a mathematical theory, what we learn to do is to make use of the statements of that language: we learn when they may be established by computation, and how to carry out the relevant computations, we learn from what they may be inferred and what may be inferred from them, that is, what rôle they play in mathematical proofs and how they can be applied in extra-mathematical contexts, and perhaps we learn also what plausible arguments can render them probable. These things are all that we are shown when we are learning the meanings of the expressions of the language of the mathematical theory in question, because they are all that we can be shown: and, likewise, our proficiency in making the correct use of the statements and expressions of the language is all that others have from which to judge whether or not we have acquired a grasp of their meanings. Hence it can only be in the capacity to make a correct use of the statements of the language that a grasp of their meanings, and those of the symbols and expressions which they contain, can consist. To suppose that there is an ingredient of meaning which transcends the use that is made of that which carries the meaning is to suppose that someone might have learned all that is directly taught when the language of a mathematical theory is taught to him, and might then behave in every way like someone who understood the language, and yet not actually understand it, or understand it only incorrectly. But to suppose this is to make meaning ineffable, that is, in principle incommunicable. If this is possible, then no one individual ever has a guarantee that he is understood by any other individual; for all he knows, or can ever know, everyone else may attach to his words or to the symbols which he employs a meaning quite different fom that which he attaches to them. A notion of meaning so private to the individual is one that has become completely irrelevant to mathematics as it is actually practised, namely as a body of theory on which many individuals are corporately engaged, an enquiry within which each can communicate his results to others.

It might seem that an approach to meaning which regarded it as exhaustively determined by use would rule out any form of revisionism. If use constitutes meaning, then, it might seem, use is beyond criticism: there can be no place for rejecting any established mathematical practice, such as the use of certain forms of argument or modes of proof, since that practice, together with all others which are generally accepted, is simply constitutive of the meanings of our mathematical statements, and we surely have the right to make our statements mean whatever we choose that they shall mean. Such an attitude is one possible development of the thesis that use exhaustively determines meaning: it is, however, one which can, ultimately, be supported only by the adoption of a holistic view of language. On such a view, it is illegitimate to ask after the content of any single statement, or even after that of any one theory, say a mathematical or a physical theory; the significance of each statement or of each deductively systematised body of statements is modified by the multiple connections which it has, direct and remote, with other statements in other areas of our language taken as a whole, and so there is no adequate way of understanding the statement short of knowing the entire language. Or, rather, even this image is false to the facts: it is not that a statement or even a theory has, as it were, a primal meaning which then gets modified by the interconnections that are established with other statements and other theories; rather, its meaning simply consists in the place which it occupies in the complicated network which constitutes the totality of our linguistic practices. The only thing to which a definite content may be attributed is the totality of all that we are, at a given time, prepared to assert; and there can be no simple model of the content which that totality of assertions embodies; nothing short of a complete knowledge of the language can reveal it.

Frequently such a holistic view is modified to the extent of admitting a class of observation statements which can be regarded as more or less directly registering our immediate experience, and hence as each carrying a determinate individual content. These observation statements lie, in Quine's famous image of language, at the periphery of the articulated structure formed by all the sentences of our language, where alone experience impinges. To these peripheral sentences, meanings may be ascribed in a more or less straightforward manner, in terms of the observational stimuli which prompt assent to and dissent from them. No comparable model of meaning is available for the sentences which lie further towards the interior of the structure: an understanding of them consists solely in a grasp of their place in the structure as a whole and their interaction with its other constituent sentences. Thus, on such a view, we may accept a mathematical theory, and admit its theorems as true, only because we find in practice that it serves as a convenient substructure deep in the interior of the complex structure which forms the total theory: there can be no question of giving a representation of the truth-conditions of the statements of the mathematical theory under which they may be judged individually as acceptable, or otherwise, in isolation from the rest of language.

Such a conception bears an evident analogy with Hilbert's view of classical mathematics; or, more accurately, with Boole's view of his logical calculus. For Hilbert, a definite individual content, according to which they may be individually judged as correct or incorrect, may legitimately be ascribed only to a very narrow range of statements of elementary number theory: these correspond to the observation statements of the holistic conception of language. All other statements of mathematics are devoid of such a content, and serve only as auxiliaries, though psychologically indispensable auxiliaries, to the recognition as correct of the finitistic statements which alone are individually meaningful. The other mathematical statements are not, on such a view, devoid of significance: but their significance lies wholly in the rôle which they play within the mathematical theories to which they belong, and which are themselves significant precisely because they enable us to establish the correctness of finitistic statements. Boole likewise distinguished, amongst the formulas of his logical calculus, those which were interpretable from those which were uninterpretable: a deduction might lead from some interpretable formulas as premisses, via uninterpretable formulas as intermediate steps, to a conclusion which was once more interpretable.

The immediately obvious difficulty about such a manner of construing a mathematical, or any other, theory is to know how it can be justified. How can we be sure that the statements or formulas to which we ascribe a content, and which are derived by such a means, are true? The difference between Hilbert and Boole, in this respect, was that Hilbert took the demand for justification seriously, and saw the business of answering it as the prime task for his philosophy of mathematics, while Boole simply ignored the question. Of course, the most obvious way to find a justification is to extend the interpretation to all the statements or formulas with which we are concerned, and, in the case of Boole's calculus, this is very readily done, and indeed yields a great simplification of the calculus. Even in Hilbert's case, the consistency proof, once found, does yield an interpretation of the infinitistic statements, though one which is relative to the particular proof in which they occur, not one uniform for all contexts. Without such a justification, the operation of the mechanism of the theory or the language remains quite opaque to us; and it is because the holist is oblivious of the demand for justification, or of the unease which the lack of one causes us, that I said that he is to be compared to Boole rather than to Hilbert. In his case, the question would become: With what right do we feel an assurance that the observation statements deduced with the help of the complex theories, mathematical, scientific and otherwise, embedded in the interior of the total linguistic structure, are true, when these observation statements are interpreted in terms of their stimulus meanings? To this the holist attempts no answer, save a generalised appeal to induction: these theories have 'worked' in the past, in the sense of having for the most part yielded true observation statements, and so we have confidence that they will continue to work in the future.

The path of thought which leads from the thesis that use exhaustively determines meaning to an acceptance of intuitionistic logic as the correct logic for mathematics is one which rejects a holistic view of mathematics and insists that each statement of any mathematical theory must have a determinate individual content. A grasp of this content cannot, in general, consist of a piece of verbalisable knowledge, but must be capable of being fully manifested by the use of the statement: but that does not imply that every aspect of its existing use is sacrosanct. An existing practice in the use of a certain fragment of language is capable of being subjected to criticism if it is impossible to systematise it, that is, to frame a model whereby each sentence carries a determinate content which can, in turn, be explained in terms of the use of that sentence. What makes it possible that such a practice may prove to be incoherent and therefore in need of revision is that there are different aspects to the use of a sentence; if the whole practice is to be capable of systematisation in the present sense, there must be a certain harmony between these different aspects. This is already apparent from the holistic examples already cited. One aspect of the use of observation statements lies in the propensities we have acquired to assent to and dissent from them under certain types of stimuli; another lies in the possibility of deducing them by means of non-observational statements, including highly theoretical ones. If the linguistic system as a whole is to be coherent, there must be harmony between these two aspects: it must not be possible to deduce observation

statements from which the perceptual stimuli require dissent. Indeed, if the observation statements are to retain their status as observation statements, a stronger demand must be made: of an observation statement deduced by means of theory, it must hold that we can place ourselves in a situation in which stimuli occur which require assent to it. This condition is thus a demand that, in a certain sense, the language as a whole be a conservative extension of that fragment of the language containing only observation statements. In just the same way, Hilbert's philosophy of mathematics requires that classical number theory, or even classical analysis, be a conservative extension of finitistic number theory.

For utterances considered quite generally, the bifurcation between the two aspects of their use lies in the distinction between the conventions governing the occasions on which the utterance is appropriately made and those governing both the responses of the hearer and what the speaker commits himself to by making the utterance: schematically, between the *conditions for* the utterance and the *consequences of* it. Where, as in mathematics, the utterances with which we are concerned are *statements,* that is, utterances by means of which assertions can be effected, this becomes the distinction between the grounds on which the statement can be asserted and its inferential consequences, the conclusions that can be inferred from it. Plainly, the requirement of harmony between these in respect of some type of statement is the requirement that the addition of statements of that type to the language produces a conservative extension of the language; i.e., that it is not possible, by going via statements of this type as intermediaries, to deduce from premisses not of that type conclusions, also not of that type, which could not have been deduced before. In the case of the logical constants, a loose way of putting the requirement is to say that there must be a harmony between the introduction and elimination rules; but, of course, this is not accurate, since the whole system has to be considered (in classical logic, for example, it is possible to infer a disjunctive statement, say by double negation elimination, without appeal to the rule of disjunction introduction). An alternative way of viewing the dichotomy between the two principal aspects of the use of statements is as a contrast between *direct* and *indirect* means of establishing them. So far as a logically complex statement is concerned, the introduction rules governing the logical constants occurring in the statement display the most direct means of establishing the statement, step by step in accordance with its logical structure; but the statement may be accepted on the basis of a complicated deduction which relies also on elimination rules, and we require a harmony which obtains only if a statement that has been indirectly established always could (in some sense of 'could') have been established directly.

Here again the demand is that the admission of the more complex inferences yield a conservative extension of the language. When only introduction rules are used, the inference involves only statements of logical complexity no greater than that of the conclusion: we require that the derivation of a statement by inferences involving statements of greater logical complexity shall be possible only when its derivation by the more direct means is in some sense already possible.

On any molecular view of language – any view on which individual sentences carry a content which belongs to them in accordance with the way they are compounded out of their own constituents, independently of other sentences of the language not involving those constituents – there must be some demand for harmony between the various aspects of the use of sentences, and hence some possibility of criticising or rejecting existing practice when it does not display the required harmony. Exactly what the harmony is which is demanded depends upon the theory of meaning accepted for the language, that is, the general model of that in which the content of an individual sentence consists; that is why I rendered the above remarks vague by the insertion of phrases like 'in some sense'. It will always be legitimate to demand, of any expression or form of sentence belonging to the language, that its addition to the language should yield a conservative extension; but, in order to make the notion of a conservative extension precise, we need to appeal to some concept such as that of truth or that of being assertible or capable in principle of being established, or the like; and just which concept is to be selected, and how it is to be explained, will depend upon the theory of meaning that is adopted.

A theory of meaning, at least of the kind with which we are mostly familiar, seizes upon some one general feature of sentences (at least of assertoric sentences, which is all we need be concerned with when considering the language of mathematics) as central: the notion of the content of an individual sentence is then to be explained in terms of this central feature. The selection of some one such feature of sentences as central to the theory of meaning is what is registered by philosophical dicta of the form, 'Meaning is . . .' – e.g., 'The meaning of a sentence is the method of its verification', 'The meaning of a sentence is determined by its truth-conditions', etc. (The slogan 'Meaning is use' is, however, of a different character: the 'use' of a sentence is not, in this sense, a *single* feature; the slogan simply restricts the *kind* of feature that may legitimately be appealed to as constituting or determining meaning.) The justification for thus selecting some one single feature of sentences as central – as being that in which their individual meanings consist – is that it is hoped that every other feature of the use of sentences can be derived,

in a uniform manner, from this central one. If, e.g., the notion of truth is taken as central to the theory of meaning, then the meanings of individual expressions will consist in the manner in which they contribute to determining the truth-conditions of sentences in which they occur; but this conception of meaning will be justified only if it is possible, for an arbitrary assertoric sentence whose truth-conditions are taken as known, to describe, in terms of the notion of truth, our actual practice in the use of such a sentence; that is, to give a general characterisation of the linguistic practice of making assertions, of the conditions under which they are made and the responses which they elicit. Obviously, we are very far from being able to construct such a general theory of the use of sentences, of the practice of speaking a language; equally obviously, it is likely that, if we ever do attain such an account, it will involve a considerable modification of the ideal pattern under which the account will take a quite general form, irrespective of the individual content of the sentence as given in terms of whatever is taken as the central notion of the theory of meaning. But it is only to the extent that we shall eventually be able to approximate to such a pattern that it is possible to give substance to the claim that it is in terms of some *one* feature, such as truth or verification, that the individual meanings of sentences and of their component expressions are to be given.

It is the multiplicity of the different features of the use of sentences, and the consequent legitimacy of the demand, given a molecular view of language, for harmony between them, that makes it possible to criticise existing practice, to call in question uses that are actually made of sentences of the language. The thesis with which we started, that use exhaustively determines meaning, does not, therefore, conflict with a revisionary attitude to some aspect of language: what it does do is to restrict the selection of the feature of sentences which is to be treated as central to the theory of meaning. On a platonistic interpretation of a mathematical theory, the central notion is that of truth: a grasp of the meaning of a sentence belonging to the language of the theory consists in a knowledge of what it is for that sentence to be true. Since, in general, the sentences of the language will not be ones whose truth-value we are capable of effectively deciding, the condition for the truth of such a sentence will be one which we are not, in general, capable of recognising as obtaining whenever it obtains, or of getting ourselves into a position in which we can so recognise it. Nevertheless, on the theory of meaning which underlies platonism, an individual's grasp of the meaning of such a sentence consists in his knowledge of what the condition is which has to obtain for the sentence to be true, even though the condition is one which he cannot, in general, recognise as obtaining when it does obtain.

This conception violates the principle that use exhaustively determines meaning; or, at least, if it does not, a strong case can be put up that it does, and it is this case which constitutes the first type of ground which appears to exist for repudiating classical in favour of intuitionistic logic for mathematics. For, if the knowledge that constitutes a grasp of the meaning of a sentence has to be capable of being manifested in actual linguistic practice, it is quite obscure in what the knowledge of the condition under which a sentence is true can consist, when that condition is not one which is always capable of being recognised as obtaining. In particular cases, of course, there may be no problem, namely when the knowledge in question may be taken as verbalisable knowledge, i.e. when the speaker is able to *state,* in other words, what the condition is for the truth of the sentence; but, as we have already noted, this cannot be the general case. An ability to state the condition for the truth of a sentence is, in effect, no more than an ability to express the content of the sentence in other words. We accept such a capacity as evidence of a grasp of the meaning of the original sentence on the presumption that the speaker understands the words in which he is stating its truth-condition; but at some point it must be possible to break out of the circle: even if it were always possible to find an equivalent, understanding plainly cannot in general consist in the ability to find a synonymous expression. Thus the knowledge in which, on the platonistic view, a grasp of the meaning of a mathematical statement consists must, in general, be implicit knowledge, knowledge which does not reside in the capacity to state that which is known. But, at least on the thesis that use exhaustively determines meaning, and perhaps on any view whatever, the ascription of implicit knowledge to someone is meaningful only if he is capable, in suitable circumstances, of fully manifesting that knowledge. (Compare Wittgenstein's question why a dog cannot be said to expect that his master will come home next week.) When the sentence is one which we have a method for effectively deciding, there is again no problem: a grasp of the condition under which the sentence is true may be said to be manifested by a mastery of the decision procedure, for the individual may, by that means, get himself into a position in which he can recognise that the condition for the truth of the sentence obtains or does not obtain, and we may reasonably suppose that, in this position, he displays by his linguistic behaviour his recognition that the sentence is, respectively, true or false. But, when the sentence is one which is not in this way effectively decidable, as is the case with the vast majority of sentences of any interesting mathematical theory, the situation is different. Since the sentence is, by hypothesis, effectively undecidable, the condition which must, in general, obtain for it to be true is not one which we are capable of recognising whenever it obtains, or of getting ourselves in a position to do so. Hence any behaviour which displays a capacity for acknowledging the sentence as being true in all cases in which the condition for its truth can be recognised as obtaining will fall short of being a full manifestation of the knowledge of the condition for its truth: it shows only that the condition can be recognised in certain cases, not that we have a grasp of what, in general, it is for that condition to obtain even in those cases when we are incapable of recognising that it does. It is, in fact, plain that the knowledge which is being ascribed to one who is said to understand the sentence is knowledge which transcends the capacity to manifest that knowledge by the way in which the sentence is used. The platonistic theory of meaning cannot be a theory in which meaning is fully determined by use.

If to know the meaning of a mathematical statement is to grasp its use; if we learn the meaning by learning the use, and our knowledge of its meaning is a knowledge which we must be capable of manifesting by the use we make of it: then the notion of *truth,* considered as a feature which each mathematical statement either determinately possesses or determinately lacks, independently of our means of recognising its truth-value, cannot be the central notion for a theory of the meanings of mathematical statements. Rather, we have to look at those things which are actually features of the use which we learn to make of mathematical statements. What we actually learn to do, when we learn some part of the language of mathematics, is to recognise, for each statement, what counts as establishing that statement as true or as false. In the case of very simple statements, we learn some computation procedure which decides their truth or falsity: for more complex statements, we learn to recognise what is to be counted as a proof or a disproof of them. That is the practice of which we acquire a mastery: and it is in the mastery of that practice that our grasp of the meanings of the statements must consist. We must, therefore, replace the notion of truth, as the central notion of the theory of meaning for mathematical statements, by the notion of *proof:* a grasp of the meaning of a statement consists in a capacity to recognise a proof of it when one is presented to us, and a grasp of the meaning of any expression smaller than a sentence must consist in a knowledge of the way in which its presence in a sentence contributes to determining what is to count as a proof of that sentence. This does not mean that we are obliged uncritically to accept the canons of proof as conventionally acknowledged. On the contrary, as soon as we construe the logical constants in terms of this conception of meaning, we become aware that certain forms of reasoning which are conventionally accepted are devoid of justification. Just because the conception of meaning in terms of proof is as much a molecular, as opposed to holistic, theory of meaning as that of meaning in terms of truth-conditions, forms of inference stand in need of justification, and are open to being rejected as unjustified. Our mathematical practice has been disfigured by a false conception of what our understanding of mathematical theories consisted in.

This sketch of one possible route to an account of why, within mathematics, classical logic must be abandoned in favour of intuitionistic logic obviously leans heavily upon Wittgensteinian ideas about language. Precisely because it rests upon taking with full seriousness the view of language as an instrument of social communication, it looks very unlike traditional intuitionist accounts, which, notoriously, accord a minimum of importance to language or to symbolism as a means of transmitting thought, and are constantly disposed to slide in the direction of solipsism. However, I said at the outset that my concern in this paper was not in the least with the exegesis of actual intuitionist writings: however little it may jibe with the view of the intuitionists themselves, the considerations that I have sketched appear to me to form one possible type of argument in favour of adopting an intuitionistic version of mathematics in place of a classical one (at least as far as the logic employed is concerned), and, moreover, an argument of considerable power. I shall not take the time here to attempt an evaluation of the argument, which would necessitate enquiring how the platonist might reply to it, and how the debate between them would then proceed: my interest lies, rather, in asking whether this is the only legitimate route to the adoption of an intuitionistic logic for mathematics.

Now the first thing that ought to strike us about the form of argument which I have sketched is that it is virtually independent of any considerations relating specifically to the *mathematical* character of the statements under discussion. The argument involved only certain considerations within the theory of meaning of a high degree of generality, and could, therefore, just as well have been applied to any statements whatever, in whatever area of language. The argument told in favour of replacing, as the central notion for the theory of meaning, the condition under which a statement is true, whether we know or can know when that condition obtains, by the condition under which we acknowledge the statement as conclusively established, a condition which we must, by the nature of the case, be capable of effectively recognising whenever it obtains. Since we were concerned with mathematical statements, which we recognise as true by means of a proof (or, in simple cases, a computation), this meant replacing the notion of truth by that of proof: evidently, the appropriate generalisation of this, for statements of an arbitrary kind, would be the

replacement of the notion of truth, as the central notion of the theory of meaning, by that of verification; to know the meaning of a statement is, on such a view, to be capable of recognising whatever counts as verifying the statement, i.e. as conclusively establishing it as true. Here, of course, the verification would not ordinarily consist in the bare occurrence of some sequence of sense-experiences, as on the positivist conception of the verification of a statement. In the mathematical case, that which establishes a statement as true is the production of a deductive argument terminating in that statement as conclusion; in the general case, a statement will, in general, also be established as true by a process of reasoning, though here the reasoning will not usually be purely deductive in character, and the premisses of the argument will be based on observation; only for a restricted class of statements – the observation statements – will their verification be of a purely observational kind, without the mediation of any chain of reasoning or any other mental, linguistic or symbolic process.

It follows that, in so far as an intuitionist position in the philosophy of mathematics (or, at least, the acceptance of an intuitionistic logic for mathematics) is supported by an argument of this first type, similar, though not necessarily identical, revisions must be made in the logic accepted for statements of other kinds. What is involved is a thesis in the theory of meaning of the highest possible level of generality. Such a thesis is vulnerable in many places: if it should prove that it cannot be coherently applied to any one region of discourse, to any one class of statements, then the thesis cannot be generally true, and the general argument in favour of it must be fallacious. Construed in this way, therefore, a position in the philosophy of mathematics will be capable of being undermined by considerations which have nothing directly to do with mathematics at all.

Is there, then, any alternative defence of the rejection, for mathematics, of classical in favour of intuitionistic logic? Is there any such defence which turns on the fact that we are dealing with *mathematical* statements in particular, and leaves it entirely open whether or not we wish to extend the argument to statements of any other general class?

Such a defence must start from some thesis about mathematical statements the analogue of which we are free to reject for statements of other kinds. It is plain what this thesis must be: namely the celebrated thesis that mathematical statements do not relate to an objective mathematical reality existing independently of us. The adoption of such a view apparently leaves us free either to reject or to adopt an analogous view for statements of any other kind. For instance, if we are realists about the physical universe, then we may contrast mathematical statements with statements ascribing physical properties to material objects: on this combination of views, material-object statements do relate to an objective reality existing independently of ourselves, and are rendered true or false, independently of our knowledge of their truth-values or of our ability to attain such knowledge or the particular means, if any, by which we do so, by that independently existing reality; the assertion that mathematical statements relate to no such external reality gains its substance by contrast with the physical case. Unlike material objects, mathematical objects are, on this thesis, creations of the human mind: they are objects of thought, not merely in the sense that they can be thought about, but in the sense that their being is to be thought of; for them, *esse est concipi*.

On such a view, a conception of meaning as determined by truth-conditions is available for any statements which do relate to an independently existing reality, for then we may legitimately assume, of each such statement, that it possesses a determinate truth-value, true or false, independently of our knowledge, according as it does or does not agree with the constitution of that external reality which it is about. But, when the statements of some class do not relate to such an external reality, the supposition that each of them possesses such a determinate truth-value is empty, and we therefore cannot regard them as being given meanings by associating truth-conditions with them; we have, in such a case, *faute de mieux*, to take them as having been given meaning in a different way, namely by associating with them conditions of a different kind - conditions that we are capable of recognising when they obtain - namely, those conditions under which we take their assertion or their denial as being conclusively justified.

The first type of justification of intuitionistic logic which we considered conformed to Kreisel's dictum, 'The point is not the existence of mathematical objects, but the objectivity of mathematical truth': it bore directly upon the claim that mathematical statements possess objective truth-values, without raising the question of the ontological status of mathematical objects or the metaphysical character of mathematical reality. But a justification of the second type violates the dictum: it makes the question whether mathematical statements possess objective truth-values depend upon a prior decision as to the being of mathematical objects. And the difficulty about it lies in knowing on what we are to base the premiss that mathematical objects are the creations of human thought in advance of deciding what is the correct model for the meanings of mathematical statements or what is the correct conception of truth as relating to them. It appears that, on this view, before deciding whether a grasp of the meaning of a mathematical statement is to be considered as consisting in a knowledge of what has to be the case for it to be

true or in a capacity to recognise a proof of it when one is presented, we have first to resolve the metaphysical question whether mathematical objects – natural numbers, for example – are, as on the constructivist view, creations of the human mind, or, as on the platonist view, independently existing abstract objects. And the puzzle is to know on what basis we could possibly resolve this metaphysical question, at a stage at which we do not even know what model to use for our understanding of mathematical statements. We are, after all, being asked to choose between two metaphors, two pictures. The platonist metaphor assimilates mathematical enquiry to the investigations of the astronomer: mathematical structures, like galaxies, exist, independently of us, in a realm of reality which we do not inhabit but which those of us who have the skill are capable of observing and reporting on. The constructivist metaphor assimilates mathematical activity to that of the artificer fashioning objects in accordance with the creative power of his imagination. Neither metaphor seems, at first sight, especially apt, nor one more apt than the other: the activities of the mathematician seem strikingly unlike those either of the astronomer or of the artist. What basis can exist for deciding which metaphor is to be preferred? How are we to know in which respects the metaphors are to be taken seriously, how the pictures are to be used?

Preliminary reflection suggests that the metaphysical question ought not to be answered first: we cannot, as the second type of approach would have us do, *first* decide the ontological status of mathematical objects, and then, with that as premiss, deduce the character of mathematical truth or the correct model of meaning for mathematical statements. Rather, we have first to decide on the correct model of meaning – either an intuitionistic one, on the basis of an argument of the first type, or a platonistic one, on the basis of some rebuttal of it; and then one or other picture of the metaphysical character of mathematical reality will force itself on us. If we have decided upon a model of the meanings of mathematical statements according to which we have to repudiate a notion of truth considered as determinately attaching, or failing to attach, to such statements independently of whether we can now, or ever will be able to, prove or disprove them, then we shall be unable to use the picture of mathematical reality as external to us and waiting to be discovered. Instead, we shall inevitably adopt the picture of that reality as being the product of our thought, or, at least, as coming into existence only as it is thought. Conversely, if we admit a notion of truth as attaching objectively to our mathematical statements independently of our knowledge, then, likewise, the picture of mathematical reality as existing, like the galaxies, independently of our observation of it will force itself on us in an equally irresistible manner. But, when we approach the matter in this way, there is no puzzle over the interpretation of these metaphors: psychologically inescapable as they may be, their non-metaphorical content will consist entirely in the two contrasting models of the meanings of mathematical statements, and the issue between them will become simply the issue as to which of these two models is correct. If, however, a view as to the ontological status of mathematical objects is to be treated as a *premiss* for deciding between the two models of meaning, then the metaphors cannot without circularity be explained solely by reference to those models; and it is obscure how else they are to be explained.

These considerations appear, at first sight, to be reinforced by reflection upon Frege's dictum, 'Only in the context of a sentence does a name stand for anything'. We cannot refer to an object save in the course of saying something about it. Hence, any thesis concerning the ontological status of objects of a given kind must be, at the same time, a thesis about what makes a statement involving reference to such objects true, in other words, a thesis about what properties an object of that kind can have. Thus, to say that fictional characters are the creations of the imagination is to say that a statement about a fictional character can be true only if it is imagined as being true, that a fictional character can have only those properties which it is part of the story that he has; to say that something is an object of sense – that for it *esse est percipi* – is to say that it has only those properties it is perceived as having: in both cases, the ontological thesis is a ground for rejecting the law of excluded middle as applied to statements about those objects. Thus we cannot separate the question of the ontological status of a class of objects from the question of the correct notion of truth for statements about those objects, i.e. of the kind of thing in virtue of which such statements are true, when they are true. This conclusion corroborates the idea that an answer to the former question cannot serve as a premiss for an answer to the latter one.

Nevertheless, the position is not so straightforward as all this would make it appear. From the possibility of an argument of the first type for the use of intuitionistic logic in mathematics, it is evident that a model of the meanings of mathematical statements in terms of proof rather than of truth need not rest upon any particular view about the ontological character of mathematical objects. There is no substantial disagreement between the two models of meaning so long as we are dealing only with decidable statements: the crucial divergence occurs when we consider ones which are not effectively decidable, and the linguistic operation which first enables us to frame effectively undecidable mathematical statements is that of quantification over infinite totalities, in the first

place over the totality of natural numbers. Now suppose someone who has, on whatever grounds, been convinced by the platonist claim that we do not create the natural numbers, and yet that reference to natural numbers is not a mere *façon de parler*, but is a genuine instance of reference to objects: he believes, with the platonist, that natural numbers are abstract objects, existing timelessly and independently of our knowledge of them. Such a person may, nevertheless, when he comes to consider the meaning of existential and universal quantification over the natural numbers, be convinced by a line of reasoning such as that which I sketched as constituting the first type of justification for replacing classical by intuitionistic logic. He may come to the conclusion that quantification over a denumerable totality cannot be construed in terms of our grasp of the conditions under which a quantified statement is true, but must, rather, be understood in terms of our ability to recognise a proof or disproof of such a statement. He will therefore reject a classical logic for number-theoretic statements in general, admitting only intuitionistically valid arguments involving them. Such a person would be accepting a platonistic view of the existence of mathematical objects (at least the objects of number theory), but rejecting a platonistic view of the objectivity of mathematical statements.

Our question is, rather, whether the opposite combination of views is possible: whether one may consistently hold that natural numbers are the creations of human thought, but yet believe that there is a notion of truth under which each number-theoretic statement is determinately either true or false, and that it is in terms of our grasp of their truth-conditions that our understanding of number-theoretic statements is to be explained. If such a combination is possible, then, it appears, there can be no route from the ontological thesis that mathematical objects are the creations of our thought to the model of the meanings of mathematical statements which underlies the adoption of an intuitionistic logic.

This is not the only question before us: for, even if these two views cannot be consistently combined, it would not follow that the ontological thesis could serve as a premiss for the constructivist view of the meanings of mathematical statements; our difficulty was to understand how the ontological thesis could have any substance if it were not merely a picture encapsulating that conception of meaning. The answer is surely this: that, while it is surely correct that a thesis about the ontological status of objects of a given kind, e.g. natural numbers, must be understood as a thesis about that in which the truth of certain statements about those objects consists, it need not be taken as, in the first place, a thesis about the entire class of such statements; it may, instead, be understood as a thesis only about some restricted subclass of such statements, those which are basic to the very possibility of making reference to those objects. Thus, for example, the thesis that natural numbers are creations of human thought may be taken as a thesis about the sort of thing which makes a numerical equation or inequality true, or, more generally, a statement formed from such equations by the sentential operators and bounded quantification. To say that the only notion of truth we can have for number-theoretic statements generally is that which equates truth with our capacity to prove a statement is to prejudge the issue about the correct model of meaning for such statements, and therefore cannot serve as a premiss for the constructivist view of meaning. But to say that, for decidable number-theoretic statements, truth consists in provability, is not in itself to prejudge the question in what the truth of undecidable statements, involving unbounded quantification, consists: and hence the possibility is open that a view about the one might serve as a premiss for a view about the other. Our problem is to discover whether it can do so in fact: whether there is any legitimate route from the thesis that natural numbers are creations of human thought, construed as a thesis about the sort of thing which makes decidable number-theoretic statements true, to a view of the meanings of number-theoretic statements generally which would require the adoption for them of an intuitionistic rather than a classical logic.

In order to resolve this question, it is necessary for us to take a rather closer look at the notion of truth for mathematical statements, as understood intuitionistically. The most obvious suggestion that comes to mind in this connection is that the intuitionistic notion of truth conforms, just as does the classical notion, to Tarski's schema:

(T) S is true iff A,

where an instance of the schema is to be formed by replacing 'A' by some number-theoretic statement and 'S' by a canonical name of that sentence, as, e.g., in:

> 'There are infinitely many twin primes' is true iff there are infinitely many twin primes.

It is necessary to admit counter-examples to the schema (T) in any case in which we wish to hold that there exist sentences which are neither true nor false: for if we replace 'A' by such a sentence, the left-hand side of the biconditional becomes false (on the assumption that, if the negation of a sentence is true, that sentence is false), although, by hypothesis, the right-hand side is not false. But, in intuitionistic logic, that semantic principle holds good which stands to the double negation of the law of

excluded middle as the law of bivalence stands to the law of excluded middle itself: it is inconsistent to assert of any statement that it is neither true nor false; and hence there seems no obstacle to admitting the correctness of the schema (T). Of course, in doing so, we must construe the statement which appears on the right-hand side of any instance of the schema in an intuitionistic manner. Provided we do this, a truth-definition for the sentences of an intuitionistic language, say that of Heyting arithmetic, may be constructed precisely on Tarski's lines, and will yield, as a consequence, each instance of the schema (T).

However, notoriously, such an approach leaves many philosophical problems unresolved. The truth-definition tells us, for example, that

'$598017+246532=844549$' is true

just in the case in which $598017+246532=844549$. We may perform the computation, and discover that $598017+246532$ does indeed equal 844549: but does that mean that the equation was already true before the computation was performed, or that it would have been true even if the computation had never been performed? The truth-definition leaves such questions quite unanswered, because it does not provide for inflections of tense or mood of the predicate 'is true': it has been introduced only as a predicate as devoid of tense as are all ordinary mathematical predicates; but its rôle in our language does not reveal why such inflections of tense or even of mood should be forbidden.

These difficulties raise their heads as soon as we make the attempt to introduce tense into mathematics, as intuitionism provides us with some inclination to do; this can be seen from the problems surrounding the theory of the creative subject. These problems are well brought out in Troelstra's discussion of the topic. It is evident that we ought to admit as an axiom

(α) $(\vdash_n A) \rightarrow A$;

if we know that, at any stage, A has been (or will be) proved, then we are certainly entitled to assert A. But ought we to admit the converse in the form

(β) $A \rightarrow \exists n(\vdash_n A)$?

Its double negation

(γ) $A \rightarrow \neg\neg\exists n(\vdash_n A)$

is certainly acceptable: if we know that A is true, then we shall certainly never be able to assert, at least on purely mathematical grounds, that it will never be proved. But can we equate truth with the obtaining of a proof at some stage, in the past or in the future, as the equivalence:

$$(\delta) \qquad A \leftrightarrow \exists n(\vdash_n A)$$

requires us to do? (To speak of 'truth' here seems legitimate, since, while Tarski's truth-predicate is a predicate of sentences, the sentential operator to which it corresponds is a redundant one, which can be inserted before or deleted from in front of any clause without change of truth-value.)

If we accept the axiom (β), and hence the equivalence (δ), we run into certain difficulties, on which Troelstra comments. The operator '$\exists n(\vdash_n \ldots)$' becomes a redundant truth-operator, and hence may be distributed across any logical constant, as in

$$(\epsilon) \qquad (\vdash_k \forall m A(m)) \rightarrow \forall m \exists n(\vdash_n A(m)).$$

As Troelstra observes, this appears to have the consequence that, if we have once proved a universally quantified statement, we are in some way committed to producing, at some time in the future, individual proofs of all its instances, whereas, palpably, we are under no such constraint. The solution to which he inclines is that proposed by Kreisel, namely that the operator '$\vdash_n$' must be so construed that a proof, at stage n, of a universally quantified statement counts as being, at the same time, a proof of each instance, so that we could assert the stronger thesis

$$(\zeta) \qquad (\vdash_k \forall m A(m)) \rightarrow \forall m(\vdash_k A(m)).$$

(Troelstra in fact recommends this interpretation on separate grounds, as enabling us to escape a paradox about constructive functions; he himself points out, however, that this paradox can alternatively be avoided by introducing distinctions of level which seem intrinsically plausible.) The difficulty about this solution is that it must be extended to every recognised logical consequence. From

$$(\eta) \qquad (m \leqslant n \;\&\; (\vdash_m A)) \rightarrow (\vdash_n A)$$

we have

$$(\theta) \qquad (n = \max(m,k) \;\&\; (\vdash_m A) \;\&\; (\vdash_k C)) \rightarrow ((\vdash_n A) \;\&\; (\vdash_n C)),$$

while from (δ) we obtain

$$(\iota) \qquad (\vdash_m A) \;\&\; (\vdash_k (A \rightarrow B)) \rightarrow \exists n(\vdash_n B).$$

We could in the same way complain that this committed us, whenever we had proved a statement A and had recognised some other statement B as being a consequence of A, to actually drawing that consequence some

time in the future; and, if our interpretation of the operator '$\vdash_n$' is to be capable of dealing with this difficulty in the same way as with the special case of instances of a universally quantified statement, we should have to allow that a proof that a theorem had a certain consequence was, at the same time, a proof of that consequence, and, likewise, that a proof of a statement already known to have a certain consequence was, at the same time, a proof of that consequence; we should, that is, have to accept the law

$$(\kappa) \qquad (n = \max(m,k) \;\&\; (\vdash_m A) \;\&\; (\vdash_k (A \rightarrow B))) \rightarrow (\vdash_n B).$$

We should thus have so to construe the notion of proof that a proof of a statement is taken as simultaneously constituting a proof of anything that has already been recognised as a consequence of that statement. We can, no doubt, escape having to say that it is simultaneously a proof of whatever, in a platonistic sense, is as a matter of fact an intuitionistic consequence of the statement: but when are we to be said to have recognised that one statement is a consequence of another? If a proof of a universally quantified statement is simultaneously a proof of all its instances, it is difficult to see how we can avoid conceding that a demonstration of the validity of a schema of first-order predicate logic is simultaneously a demonstration of the truth of all its instances, or an acceptance of the induction schema simultaneously an acceptance of all cases of induction. The resulting notion of proof would be far removed indeed from actual mathematical experience, and could not be explained as no more than an idealisation of it.

The trouble with all this is that, as a representation of actual mathematical experience, we are operating with too simplified a notion of proof. The axiom (η) is acceptable in the sense that, prescinding from the occasional accident, once a theorem has been proved, it always remains *available* to be subsequently appealed to: but the idea that, having acknowledged the two premisses of a modus ponens, we have *thereby* recognised the truth of the conclusion, is plausible only in a case in which we are simultaneously bearing in mind the truth of the two premisses. To have once proved a statement is not thereafter to be continuously aware of its truth: if it were, then we should indeed always know the logical consequences of everything which we know, and should have no need of proof.

Acceptance of axiom (β) leads to the conclusion that we shall eventually prove every logical consequence of everything we prove. This, as a representation of the intuitionist notion of proof, is an improvement upon Beth trees, as normally presented: for these are set up in such a way that, at any stage (node), every logical consequence of statements true at that stage is already true; the Beth trees are adapted only to situations, such as those involving free choice sequences, where new information is coming in that is not derived from the information we have at earlier stages. But the idea that we shall eventually establish every logical consequence of everything we know is implausible and arbitrary: and it cannot be rescued by construing each proof as, implicitly, a proof also of the consequences of the statement proved, save at the cost of perverting the whole conception. If we wish to do so, there seems no reason why we should not take the stages represented by the numerical subscripts as punctuated by proofs, however short the stages thereby become, and the notion of proof as relating only to what is quite explicitly proved, so that, at each stage, one and only new statement is proved, and consider what axioms hold under the resulting interpretation of the symbol '$\vdash_n$'. It thus appears that, under this interpretation, the axiom (β) must be rejected in favour of the weaker axiom (γ).

Looked at in another way, however, the stronger axiom (β) seems entirely acceptable. If, that is, we interpret the implication sign in its intuitionistic sense, the axiom merely says that, given a proof of A, we can effectively find a proof that A was proved at some stage; and this seems totally innocuous and banal. But, if axiom (β) is innocuous, how did we arrive at our earlier difficulties? The only possibility seems to be that our logical laws are themselves at fault. For instance, the law

(λ) $$\forall x A(x) \rightarrow A(m)$$

leads, via axiom (β), to the conclusion

(μ) $$\forall x A(x) \rightarrow \exists n(\vdash_n A(m)),$$

which appears, on the present interpretation of '$\vdash_n$', to say that we shall explicitly prove every instance of every universally quantified statement which we prove; so perhaps the error lies in the law (λ) itself. A law such as (λ) is ordinarily justified by saying that, given a proof of $\forall x A(x)$, we can, for each m, effectively find a proof of $A(m)$. If this is to remain a sufficient justification of (μ), then (μ) must be construed as saying that, given a proof of $\forall x A(x)$, we can effectively find a proof that $A(m)$ will be proved at some stage. How can we do this, for given m? Obviously, by proving $A(m)$ and noting the stage at which we do so. This means, then, that the existentially quantified statement

(ν) $$\exists n(\vdash_n A(m))$$

is to be so understood that its assertion does not amount to a claim that we shall, as a matter of fact, prove $A(m)$ at some stage n, but only that we are capable of bringing it about that $A(m)$ is proved at some stage.

Our difficulties thus appear to have arisen from understanding the existential quantifier in (β) in an excessively classical or realistic manner, namely as meaning that there will in fact be a stage n at which the statement is proved, rather than as meaning that we have an effective means, if we choose to apply it, of making it the case that there is such a stage. The point here is that it is not merely a question of interpreting the existential quantifier intuitionistically rather than classically in the sense that we can assert that there is a stage n at which a statement will be proved only if we have an effective means for identifying a particular such stage. Rather, if quantification over temporal stages is to be introduced into mathematical statements, then it must be treated like quantification over mathematical objects and mathematical constructions: the assertion that there is a stage n at which such-and-such will hold is justified provided that we possess an enduring capability of bringing about such a stage, regardless of whether we ever exercise this capability or not.

The confusions concerning the theory of the creative subject which we have been engaged in disentangling arose in part from a perfectly legitimate desire, to relate the intuitionistic truth of a mathematical statement with a use of the logical constants which is alien to intuitionistic mathematics. Troelstra's difficulties sprang from his desire to construe the expression '$\exists n(\vdash_n A)$' as meaning that A would in fact be proved at some stage: but, whether we interpret the existential quantifier classically or constructively, such a way of construing it fails to jibe with the way it and the other logical constants are construed within ordinary mathematical statements, and hence, however we try to modify our notion of a statement's being proved, we shall not obtain anything equivalent to the mathematical statement A itself. Nevertheless, the desire to express the condition for the intuitionistic truth of a mathematical statement in terms which do not presuppose an understanding of the intuitionistic logical constants as used within mathematical statements is entirely licit. Indeed, if it were impossible to do so, intuitionists would have no way of conveying to platonist mathematicians what it was that they were about: we should have a situation quite different from that which in fact obtains, namely one in which some people found it natural to extend basic computational mathematics in a classical direction, and others found it natural to extend it in an intuitionistic direction, and neither could gain a glimmering of what the other was at. That we are not in this situation is because intuitionists and platonists can find a common ground, namely statements, both mathematical and non-mathematical, which are, in the view of both, decidable, and about whose meaning there is therefore no serious dispute and which both sides agree obey a classical logic. Each party can, accordingly, by use of and reference to these unproblematic statements, explain to the other what his conception of meaning is for those mathematical statements which are in dispute. Such an explanation may not be accepted as legitimate by the other side (the whole point of the intuitionist position is that undecidable mathematical statements cannot legitimately be given a meaning by laying down truth-conditions for them in the platonistic manner): but at least the conception of meaning held by each party is not wholly opaque to the other.

This dispute between platonists and intuitionists is a dispute over whether or not a realist interpretation is legitimate for mathematical statements: and the situation I have just indicated is quite characteristic for disputes concerning the legitimacy of a realist interpretation of some class of statements, and is what allows a *dispute* to take place at all. Typically, in such a dispute there is some auxiliary class of statements about which both sides agree that a realist interpretation is possible (depending upon the grounds offered by the anti-realists for rejecting a realist interpretation for statements of the disputed class, this auxiliary class may or may not consist of statements agreed to be effectively decidable); and, typically, it is in terms of the truth-conditions of statements of this auxiliary class that the anti-realist frames his conception of meaning, his non-classical notion of truth, for statements of the disputed class, while the realist very often appeals to statements of the auxiliary class as providing an analogy for his conception of meaning for statements of the disputed class. Thus, when the dispute concerns statements about the future, statements about the present will form the auxiliary class; when it concerns statements about material objects, the auxiliary class will consist of sense-data statements; when the dispute concerns statements about character-traits, the auxiliary class will consist of statements about actual or hypothetical behaviour; and so on.

If the intuitionistic notion of truth for mathematical statements can be explained only by a Tarski-type truth-definition which takes for granted the meanings of the intuitionistic logical constants, then the intuitionist notion of truth, and hence of meaning, cannot be so much as conveyed to anyone who does not accept it already, and no debate between intuitionists and platonists is possible, because they cannot communicate with one another. It is therefore wholly legitimate, and, indeed, essential, to frame the condition for the intuitionistic truth of a mathematical statement in terms which are intelligible to a platonist and do not beg any questions, because they employ only notions which are not in dispute.

The obvious way to do this is to say that a mathematical statement is intuitionistically true if there exists an (intuitionistic) proof of it, where the existence of a proof does not consist in its platonic existence in a

realm outside space and time, but in our actual possession of it. Such a notion of truth, obvious as it is, already departs at once from that supplied by the analogue of the Tarski-type truth-definition, since the predicate 'is true', thus explained, is significantly tensed: a statement not now true may later become true. For this reason, when 'true' is so construed, the schema (T) is incorrect: for the negation of the right-hand side of any instance will be a mathematical statement, while the negation of the left-hand side will be a non-mathematical statement, to the effect that we do not as yet possess a proof of a certain mathematical statement, and hence the two sides cannot be equivalent. We might, indeed, seek to restore the equivalence by replacing 'is true' on the left-hand side by 'is or will be true': but this would lead us back into the difficulties we encountered with the theory of the creative subject, and I shall not further explore it.

What does require exploration is the notion of proof being appealed to, and that also of the existence of a proof. It has often, and, I think, correctly, been held that the notion of proof needs to be specialised if it is to supply a non-circular account of the meanings of the intuitionistic logical constants. It is possible to see this by considering disjunction and existential quantification. The standard explanation of disjunction is that a construction is a proof of $A \vee B$ just in case it is a proof either of A or of B. Despite this, it is not normally considered legitimate to assert a disjunction, say in the course of a proof, only when we actually have a proof of one or other disjunct. For instance, it would be quite in order to assert that

$$10^{10^{10}}+1 \text{ is either prime or composite}$$

without being able to say which alternative held good, and to derive some theorem by means of an argument by cases. What makes this legitimate, on the standard intuitionist view, is that we have a method which is in principle effective for deciding which of the two alternatives is correct: if we were to take the trouble to apply this method, the appeal to an argument by cases could be dispensed with. Generally speaking, therefore, if we take a statement as being true only when we actually possess a proof of it, an assertion of a disjunctive statement will not amount to a claim that it is true, but only to a claim that we have a means, effective in principle, for obtaining a proof of it. This means, however, that we have to distinguish between a proof proper, a proof in the sense of 'proof' used in the explanations of the logical constants, and a cogent argument. In the course of a cogent argument for the assertibility of a mathematical statement, a disjunction of which we do not possess, an actual proof may be asserted, and an argument by cases based upon this disjunction. This argument will not itself be a proof, since any initial segment of a proof must again be a proof: it merely indicates an effective method by which we might obtain a proof of the theorem if we cared to apply it. We thus appear to require a distinction between a proof proper – a canonical proof – and the sort of argument which will normally appear in a mathematical article or textbook, an argument which we may call a 'demonstration'. A demonstration is just as cogent a ground for the assertion of its conclusion as is a canonical proof, and is related to it in this way: that a demonstration of a proposition provides an effective means for finding a canonical proof. But it is in terms of the notion of a canonical proof that the meanings of the logical constants are given. Exactly similar remarks apply to the existential quantifier.

There is some awkwardness about this way of looking at disjunction and existential quantification, namely in the divorce between the notions of truth and of assertibility. It might be replied that the significance of the act of assertion is not, in general, uniquely determined by the notion of truth: for instance, even when we take the notion of truth for mathematical statements as given, it still needs to be stipulated whether the assertion of a mathematical statement amounts to a claim to have a proof of it, or whether it may legitimately be based on what Polya calls a 'plausible argument' of a non-apodictic kind. (We can imagine people whose mathematics wholly resembles ours, save that they do not construe an assertion as embodying a claim to have more than a plausible argument.) It nevertheless remains that, if the truth of a mathematical statement consists in our possession of a canonical proof of it, while its assertion need be based on possession of no more than a demonstration, we are forced to embrace the awkward conclusion that it may be legitimate to assert a statement even though it is *known* not to be true. Still, if the sign of disjunction and the existential quantifier were the only logical constants whose explanation appeared to call for a distinction between canonical proofs and demonstrations, the distinction might be avoided altogether by modifying their explanations, to allow that a proof of a disjunction consisted in any construction of which we could recognise that it would effectively yield a proof of one or other disjunct, and similarly for existential quantification: we should then be able to say that a statement could be asserted only when it was (known to be) true.

However, the distinction is unavoidable if the explanations of universal quantification, implication and negation are to escape circularity. The standard explanation of implication is that a proof of $A \rightarrow B$ is a construction of which we can recognise that, applied to any proof of A, it would yield a proof of B. It is plain that the notion of proof being used here cannot be one which admits unrestricted use of modus ponens: for,

if it did, the explanation would be quite empty. We could admit anything we liked as constituting a proof of $A \rightarrow B$, and it would remain the case that, given such a proof, we had an effective method of converting any proof of A into a proof of B, namely by adding the proof of $A \rightarrow B$ and performing a single inference by modus ponens. Obviously, this is not what is intended: what is intended is that the proof of $A \rightarrow B$ should supply a means of converting a proof of A into a proof of B without appeal to modus ponens, at least, without appeal to any modus ponens containing $A \rightarrow B$ as a premiss. The kind of proof in terms of which the explanation of implication is being given is, therefore, one of a restricted kind. On the assumption that we have, or can effectively obtain, a proof of $A \rightarrow B$ of this restricted kind, an inference from $A \rightarrow B$ by modus ponens is justified, because it is in principle unnecessary. The same must, by parity of reasoning, hold good for any other application of modus ponens in the main (though not in any subordinate) deduction of any proof. Thus, if the intuitionistic explanation of implication is to escape, not merely circularity, but total vacuousness, there must be a restricted type of proof – canonical proof – in terms of which the explanation is given, and which does not admit modus ponens save in subordinate deductions. Arguments employing modus ponens will be perfectly valid and compelling, but they will, again, not be proofs in this restricted sense: they will be demonstrations, related to canonical proofs as supplying a means effective in principle for finding canonical proofs. Exactly similar remarks apply to universal quantification *vis-à-vis* universal instantiation and to negation *vis-à-vis* the rule ex falso quodlibet: the explanations of these operators presuppose a restricted type of proof in which the corresponding elimination rules do not occur within the main deduction.

What exactly the notion of a canonical proof amounts to is obscure. The deletion of elimination rules from a canonical proof suggests a comparison with the notion of a normalised deduction. On the other hand, Brouwer's celebrated remarks about fully analysed proofs in connection with the bar theorem do not suggest that such a proof is one from which unnecessary detours have been cut out – the proof of the bar theorem consists in great part in cutting out such detours from a proof taken already to be in 'fully analysed' form. Rather, Brouwer's idea appears to be that, in a fully analysed proof, all operations on which the proof depends will actually have been carried out. That is why such a proof may be an infinite structure: a proof of a universally quantified statement will be an operation which, applied to each natural number, will yield a proof of the corresponding instance; and, if this operation is carried out for each natural number, we shall have proofs of denumerably many statements. The conception of the mental construction which is the fully analysed proof as being an infinite structure must, of course, be interpreted in the light of the intuitionist view that all infinity is potential infinity: the mental construction consists of a grasp of general principles according to which any finite segment of the proof could be explicitly constructed. The direction of analysis runs counter to the direction of deduction; while one could not be convinced by an actually infinite proof-structure (because one would never reach the conclusion), one may be convinced by a potentially infinite one, because its infinity consists in our grasp of the principles governing its analysis. Indeed, it might reasonably be said that the standard intuitionistic meanings of the universal and conditional quantifiers involve that a proof is such a potentially infinite structure. Nevertheless, the notion of a fully analysed proof, that is, of the result of applying every operation involved in the proof, is far from clear, because it is obscure what the effect of the analysis would be on conditionals and negative statements. We can systematically display the results of applying the operation which constitutes a proof of a statement involving universal quantification over the natural numbers, because we can generate each natural number in sequence. But the corresponding application of the operation which constitutes the proof of a statement of the form $A \rightarrow B$ would consist in running through all putative canonical proofs of A and either showing, in each case, that it was not a proof of A, or transforming it into a proof of B; and, at least without a firm grasp upon the notion of a canonical proof, we have no idea how to generate all the possible candidates for being a proof of A.

The notion of canonical proof thus lies in some obscurity; and this state of affairs is not indefinitely tolerable, because, unless it is possible to find a coherent and relatively sharp explanation of the notion, the viability of the intuitionist explanations of the logical constants must remain in doubt. But, for present purposes, it does not matter just how the notion of canonical proof is to be explained; all that matters is that we require some distinction between canonical proofs and demonstrations, related to one another in the way that has been stated. Granted that such a distinction is necessary, there is no motivation for refusing to apply it to the case of disjunctions and existential statements.

Let us now ask whether we want the intuitionistic truth of a mathematical statement to consist in the existence of a canonical proof or of a demonstration. If by the 'existence' of a proof or demonstration we mean that we have actually explicitly carried one out, then either choice leaves us with certain counter-intuitive consequences. On either view, naturally, a valid rule of inference will not always lead from true prem-

isses to a true conclusion, namely if we have not explicitly drawn the inference: this will always be so on any view which equates truth with our actual possession of some kind of proof. If we take the stricter line, and hold a statement to be true only when we possess a canonical proof of it, then, as we have seen, we shall have to allow that a statement may be asserted even though it is known not to be true. If, on the other hand, we allow that a statement is true when we possess merely a demonstration of it, then truth will not distribute over disjunction: we may possess a demonstration of $A \vee B$ without having a demonstration either of A or of B. Now, admittedly, once we have admitted a significant tense for the predicate 'is true', then, as we have noted, the schema (T) cannot be maintained as in all cases correct: but our instinct is to permit as little divergence from it as possible, and it is for this reason that we are uneasy about a notion of truth which is not distributive over disjunction or existential quantification.

A natural emendation is to relax slightly the requirement that a proof or demonstration should have been explicitly given. The question is how far we may consistently go along this path. If we say merely that a mathematical statement is true just in case we are aware that we have an effective means of obtaining a canonical proof of it, this will not be significantly different from equating truth with our actual possession of a demonstration. It might be allowed that there would be some cases when we had demonstrated the premisses of, say, an inference by modus ponens in which we were aware that we could draw the conclusion, though we had not quite explicitly done so; but there will naturally be others in which we were not aware of this, i.e. had not noticed it; if it were not so, we could never discover new demonstrations. It is therefore tempting to go one step further, and say that a statement is true provided that we are in fact in possession of a means of obtaining a canonical proof of it, whether or not we are aware of the fact. Would such a step be a betrayal of intuitionist principles?

In which cases would it be correct to say that we possess an effective means of finding a canonical proof of a statement, although we do not know that we have such a means? Unless we are to suppose that we can attain so sharp a notion of a canonical proof that it would be possible to enumerate effectively all putative such proofs of a given statement (the supposition whose implausibility causes our difficulty over the notion of a fully analysed proof), there is only one such case: that in which we possess a demonstration of a disjunctive or existential statement. Such a demonstration provides us with what we recognise as an effective means (in principle) for finding a canonical proof of the disjunctive or existential statement demonstrated. Such a canonical proof, when found, will be a proof of one or other disjunct, or of one instance of the existentially quantified statement: but we cannot, in general, tell which. For example, when $A(x)$ is a decidable predicate, the decision procedure constitutes a demonstration of the disjunction '$A(\bar{n}) \vee \neg A(\bar{n})$', for specific n; but, until we apply the procedure, we do not know which of the two disjuncts we can prove. It is very difficult for us to resist the temptation to suppose that there is already, unknown to us, a determinate answer to the question which of the two disjuncts we should obtain a proof of, were we to apply the decision procedure; that, for example, that it is already the case either that, if we were to test it out, we should find that $10^{10^{10}}+1$ is prime, or that, if we were to test it out, we should find that it was composite. What is involved here is the passage from a subjunctive conditional of the form:

$$A \rightarrow (B \vee C)$$

to a disjunction of subjunctive conditionals of the form

$$(A \rightarrow B) \vee (A \rightarrow C).$$

Where the conditional is interpreted intuitionistically, this transition is, of course, invalid: but the subjunctive conditional of natural language does not coincide with the conditional of intuitionistic mathematics. It is, indeed, the case that the transition is not in general valid for the subjunctive conditional of natural language either: but, when we reflect on the cases in which the inference fails, it is difficult to avoid thinking that the present case is not one of them.

There are two obvious kinds of counter-example to this form of inference for ordinary subjunctive conditionals: perhaps they are really two sub-varieties of a single type. One is the case in which the antecedent A requires supplementation before it will yield a determinate one of the disjuncts B and C. For instance, we may safely agree that, if Fidel Castro were to meet President Carter, he would either insult him or speak politely to him; but it might not be determinately true, of either of those things, that he would do it, since it might depend upon some so far unspecified further condition, such as whether the meeting took place in Cuba or outside. Schematically, this kind of case is one in which we can assert:

$$A \rightarrow (B \vee C),$$
$$(A \,\&\, Q) \rightarrow B,$$
$$(A \,\&\, \neg Q) \rightarrow C,$$

but in which the subjunctive antecedent A neither implies nor presupposes either Q or its negation; in such a case, we cannot assert either

$A \rightarrow B$ or $A \rightarrow C$. The other kind of counter-example is that in which we do not consider the disjuncts to be determined by anything at all: no supplementation of the antecedent would be sufficient to decide between them in advance. If that light-beam were to fall upon an atom, either it would assume a higher energy level, or it would remain in its ground state; but nothing can determine for certain in advance which would happen. Similar cases will arise, for those who believe in free will in the traditional sense, in respect of human actions.

If we were to carry out the decision procedure for determining the primality or otherwise of some specific large number N, we should either obtain the result that N is prime or obtain the result that N is composite. Is this, or is it not, a case in which we may conclude that it either holds good that, if we were to carry out the procedure, we should find that N is prime, or that, if we were to carry out the procedure, we should find that N is composite? The difficulty of resisting the conclusion that it is such a case stems from the fact that it does not display either of the characteristics found in the two readily admitted types of counter-example to the form of inference we are considering. No further circumstance could be relevant to the result of the procedure – this is part of what is meant by calling it a computation; and, since at each step the outcome of the procedure is determined, how can we deny that the overall outcome is determinate also?

If we yield to this line of thought, then we must hold that every statement formed by applying a decidable predicate to a specific natural number already has a definite truth-value, true or false, although we may not know it. And, if we hold this, it makes no difference whether we chose at the outset to say that natural numbers are creations of the human mind or that they are eternally existing abstract objects. Whichever we say, our decision how to interpret undecidable statements of number theory, and, in the first place, statements of the forms $\forall x A(x)$ and $\exists x A(x)$, where $A(x)$ is decidable, will be independent of our view about the ontological status of natural numbers. For, on this view of the truth of mathematical statements, each decidable number-theoretic statement will already be determinately true or false, independently of our knowledge, just as it is on a platonistic view; any thesis about the ontological character of natural numbers will then be quite irrelevant to the interpretation of the quantifiers. As we noted, it would be possible for someone to be prepared to regard natural numbers as timeless abstract objects, and to regard decidable predicates as being determinately true or false of them, and yet to be convinced by an argument of the first type, based on quite general considerations concerning meaning, that unbounded quantification over natural numbers was not an operation which in all cases preserved the property of possessing a determinate truth-value, and therefore to fall back upon a constructivist interpretation of it. Conversely, if someone who thought of the natural numbers as creations of human thought also believed, for the reasons just indicated, that each decidable predicate was determinately true or false of each of them, he might accept a classical interpretation of the quantifiers. He would do so if he was unconvinced by the general considerations about meaning which we reviewed, i.e., by the first type of argument for the adoption of an intuitionistic logic for mathematics: the fact that he was prepared to concede that the natural numbers come into existence only in virtue of our thinking about them would play no part in his reflections on the meanings of the quantifiers. Dedekind, who declared that mathematical structures are free creations of the human mind, but nevertheless appears to have construed statements about them in a wholly platonistic manner, may perhaps be an instance of just such a combination of ideas.

One who rejects the idea that there is already a determinate outcome for the application, to any specific case, of an effective procedure is, however, in a completely different position. If someone holds that the only acceptable sense in which a mathematical statement, even one that is effectively decidable, can be said to be true is that in which this means that we presently possess an actual proof or demonstration of it, then a classical interpretation of unbounded quantification over the natural numbers is simply unavailable to him. As is frequently remarked, the classical or platonistic coneption is that such quantification represents an infinite conjunction or disjunction: the truth-value of the quantified statement is determined as the infinite sum or product of the truth-values of the denumerably many instances. Whether nor not this be regarded as an acceptable means of determining the meaning of these operators, the explanation presupposes that all the instances of the quantified statement themselves already possess determinate truth-values: if they do not, it is impossible to take the infinite sum or product of these. But if, for example, we do not hold that such a predicate as 'x is odd $\rightarrow x$ is not perfect' already has a determinate application to each natural number, though we do not know it, then it is just not open to us to think that, by attaching a quantifier to this predicate, we obtain a statement that is determinately true or false.

One question which we asked earlier was this: Can the thesis that natural numbers are creations of human thought be taken as a premiss for the adoption of an intuitionistic logic for number-theoretic statements? And another question was: What content can be given to the thesis that natural numbers are creations of human thought that does not prejudge the question what is the correct notion of truth for number-

theoretic statements in general? The tentative answer which we gave to this latter question was that the thesis might be taken as relating to the appropriate notion of truth for a restricted class of number-theoretic statements, say numerical equations, or, more generally, decidable statements. From what we have said about the intuitionistic notion of truth for mathematical statements, it has now become apparent that there is one way in which the thesis that natural numbers are creations of the human mind might be taken, namely as relating precisely to the appropriate notion of truth for decidable statements of arithmetic, which would provide a ground for rejecting a platonistic interpretation of number-theoretic statements generally, without appeal to any general thesis concerning the notion of meaning. This way of taking the thesis would amount to holding that there is no notion of truth applicable even to numerical equations save that in which a statement is true when we have actually performed a computation (or effected a proof) which justifies that statement. Such a claim must rest, as we have seen, on the most resolute scepticism concerning subjunctive conditionals: it must deny that there exists any proposition which is now true about what the result of a computation which has not yet been performed would be if it were to be performed. Anyone who can hang on to a view as hard-headed as this has no temptation at all to accept a platonistic view of number-theoretic statements involving unbounded quantification: he has a rationale for an intuitionistic interpretation of them which rests upon considerations relating solely to mathematics, and demanding no extension to other realms of discourse (save in so far as the subjunctive conditional is involved in explanations of the meanings of statements in these other realms). But, for anyone who is not prepared to be quite as hard-headed as that, the route to a defence of an intuitionistic interpretation of mathematical statements which begins from the ontological status of mathematical objects is closed; the only path that he can take to this goal is that which I sketched at the outset: one turning on the answers given to general questions in the theory of meaning.

The concept of number

GOTTLOB FREGE

Each individual number is an independent object

55. Having recognized that a statement of number is an assertion about a concept, we can attempt to supplement the leibnizian definitions of the individual numbers by means of the definitions of 0 and of 1.

Right away we might say: the number 0 applies to a concept, if no object falls under that concept. Here, however, "no" appears to have been substituted for 0, with which it is synonymous. Therefore the following wording is preferable: the number 0 applies to a concept if, no matter what *a* might be, the statement always holds that *a* does not fall under this concept.

Similarly we could say: the number 1 applies to a concept *F* if it is not the case that no matter what *a* is, *a* does not fall under *F*, and if from the statement

> '*a* falls under *F*' and '*b* falls under *F*'

it always follows that *a* and *b* are the same.

We must still define in general the transition from one number to the next. We will try the following formulation: the number $(n+1)$ applies to the concept *F* if there is an object *a* which falls under *F* and such that the number *n* applies to the concept "falling under *F* but not [identical with] *a*."

56. These definitions appear so natural, following our previous results, that an explanation is called for to show why they cannot satisfy us.

The last definition will most quickly arouse hesitation, for, strictly speaking, the sense of the expression 'the number *n* applies to the concept *G*' is just as unknown to us as that of the expression 'the number $(n+1)$ applies to the concept *F*'. To be sure, we can say by means of this and the next-to-last definition what

> 'the number $1+1$ applies to the concept *F*'

Translated by Michael S. Mahoney from Gottlob Frege, *Die Grundlagen der Arithmetik* (Breslau: 1884), pp. 67–104, 115–19.

means, and then, using this, indicate the sense of the expression

> 'the number $1+1+1$ applies to the concept *F*', etc.

But, to give a crude example, we can never decide by means of our definitions, whether the number *Julius Caesar* applies to a concept, whether this well-known conqueror of Gaul is a number or not. Furthermore, we cannot prove with the help of our attempted definitions that *a* must equal *b* if *a* applies to the concept *F* and *b* applies to the same concept. The expression '*the* number which applies to the concept *F*' would, therefore, not be justifiable, and it would consequently be completely impossible to prove a numerical equality because we could never isolate a definite number. It is only apparent that we have defined 0 and 1; as a matter of fact, we have only determined the sense of the expressions

> 'the number 0 applies to'

and

> 'the number 1 applies to';

but it is not permissible to isolate in these 0 and 1 as independent, recognizable objects.

57. Here is the place to examine somewhat more closely our statement that a statement of number involves an assertion about a concept. In the sentence 'the number 0 applies to the concept *F*', 0 is only a part of the predicate, if we consider the concept *F* as the actual subject. Therefore I have avoided calling numbers like 0, 1, 2 properties of concepts. The individual number appears as a separate independent object for the very reason that it forms only a part of the assertion. I have already called attention above to the fact that we say 'the [number] 1' and, by means of the definite article, set up 1 as an object.

This independence appears everywhere in arithmetic, e.g., in the equation '$1+1=2$'. Since the important thing here is to grasp the concept of number in such a way that it is useful for science, it needn't disturb us that in everyday usage the number appears attributively. This may always be avoided. E.g., the sentence 'Jupiter has four moons' may be rearranged to form 'The number of Jupiter's moons is four'. Here the 'is' is not to be considered merely a copula, as in the sentence 'the sky is blue'. This is shown by the fact that one can say 'the number of Jupiter's moons is four' or 'is the number four'. Here 'is' has the sense of 'is equal to', 'is the same as'. We have, therefore, an equation which asserts that the expression 'the number of Jupiter's moons' denotes the same object as the word 'four'. And the form of the equation is the reigning one in arithmetic. The fact that nothing about Jupiter or about a moon is contained in the word 'four' is no objection to this interpretation. Neither is there anything in the name 'Columbus' to suggest discovery or America, and nonetheless the same man is called both Columbus and the discoverer of America.

58. One could object that we cannot at all represent[1] to ourselves the object which we call four or the number of Jupiter's moons as something separate and indepenent. However, it is not the separateness which we have given the number that is at fault. To be sure, one would like to believe that in picturing the four spots of a die something appears which corresponds to the word 'four' – but that is an illusion. Think of a green meadow and see whether the picture changes when the indefinite article is replaced by the number 'one'. Nothing is added, but there is certainly something in the picture corresponding to the word 'green'.

If one pictures for himself the printed word 'gold', one will not at first associate any number with it. Were one now to ask himself how many letters the word has, the result would be the number 4; the picture, however, will be in no way more definite, but can remain wholly unchanged. The added concept "letter of the word 'gold'" is the very thing in which we discover the number. In the case of the four spots of a die the situation is somewhat less obvious because the concept is forced upon us so directly by the similarity of the spots that we hardly notice its intrusion. The number can be *pictured* [translator's italics] neither as a separate object nor as a property of an outward thing, because it is neither something sensible nor the property of an outward thing. The situation is probably most clear in the sense of the number 0. One will try in vain to picture 0 visible stars. To be sure, one can think of the sky completely covered up by clouds; but there is nothing in this picture which might correspond to the word 'star' or to the 0. One is only imagining a situation in which one may conclude: now no star may be seen.

59. Perhaps each word awakens some sort of picture for us, even a word like 'only'. The picture, however, need not correspond to the content of the word; it can be an entirely different one for different men. One will then probably imagine a situation which evokes a sentence in which the word occurs; or the spoken word might call forth the written word in one's memory.

This does not occur only in the case of particles. There can be no doubt that we lack any idea [picture] of our distance from the sun. For, even if

[1] In the sense of 'picture'.

we know the rule about the number of times we must multiply a unit of measure, nevertheless any attempt by this rule to sketch a picture which even slightly approaches the one desired is doomed to fail. This is, however, no reason to doubt the correctness of the computation by which the distance has been found, and it in no way hinders us in basing further conclusions on this being the distance.

60. Even such a concrete thing as the earth we cannot picture in the way that we have learned it actually to be, but rather we are satisfied with a sphere of medium size, which serves us as a symbol for the earth, knowing nevertheless that the two are very different from one another. Now although our picture often does not at all meet the requirements, still we make judgments with great certainty about an object like the earth, even where its size is concerned.

Thought often leads us far beyond the imaginable without thereby depriving us of the basis for our conclusions. Even if, as it appears, thought without mental pictures is impossible for us men, still their connection with the object of thought can be wholly superficial, arbitrary, and conventional.

The unimaginability of the content of a word is no reason, then, to deny it any meaning or to exclude it from usage. That we are nevertheless inclined to do so is probably owing to the fact that we consider words individually and ask about their meaning [in isolation], for which we then adopt a mental picture. Thus a word for which we are lacking a corresponding inner picture will seem to have no content. However, we must always consider a complete sentence. Only in [the context of] the latter do the words really have a meaning. The inner pictures which somehow sway before us (in reading the sentence) need not correspond to the logical components of the judgment. It is enough if the sentence as a whole has a sense; by means of this its parts also receive their content.

This observation seems to me to be useful in throwing light on several difficult concepts, such as that of the infinitesimal,[2] and its scope is probably not limited to mathematics.

The separateness [independence] which I require for the number is not intended to mean that a number-word used outside of the context of a sentence shall denote anything, but rather I want only to exclude its use as a predicate or attribute, for such a use somewhat alters its meaning.

[2] What is in question here is defining the sense of an equation like

$$df(x) = g(x)dx$$

rather than finding an interval bounded by two distinct points and of length dx.

61. But, one might object, even if the earth is really unimaginable, still it is an external thing having a definite place. Where, however, is the number 4? It is neither outside of us nor inside of us. Taken in spatial terms, this is correct. A determination of the place of the number 4 makes no sense. But, from this it follows only that the number 4 is not a spatial object, not that it is no object at all. Not every object is somewhere. Even our mental pictures[3] are in this sense not in us (subcutaneously). In us there are ganglia cells, blood particles, etc., but no mental pictures. Spatial predicates are not applicable to them: the one is neither right nor left of the other. Mental pictures have no distances between them which may be stated in millimeters. When nevertheless we refer to them as in us, we mean that they are subjective.

Even if the subjective has no spatial location, however, how is it possible for the number 4, which is objective, to be nowhere? Now I maintain that there is no contradiction here. The number 4 is, as a matter of fact, exactly the same for everyone who works with it; but this has nothing to do with spatiality. Not every objective object has a place.

In order to obtain concept of number, one must determine the sense of a numerical equation

62. How shall we have a number, then, if we can have no idea or picture of it? Only in the context of a sentence do words have meaning. We must, therefore, define the sense of a sentence in which a number-word occurs. This seems at first to leave a lot of latitude, but we have already determined that number-words are to be understood as standing for independent objects. This already specifies a class of sentences which must have a sense, the class of those sentences which express the recognition [of a number as the same number]. If for us the symbol a is to denote an object, then we must have a criterion which determines in every case whether b is the same as a, even if it is not always within our power to apply this criterion. In our present case, we must explain the sense of the statement:

> 'the number which applies to the concept F is the same number as that which applies to the concept G',

i.e., we must reproduce the content of this statement in another way without using the expression

> 'the number which applies to the concept F'.

In doing this, we give a general criterion for the equality of numbers.

[3]This word is understood purely psychologically, not psychophysically.

Once we have obtained such a means of grasping a definite number and recognizing it as such, we can assign it a number-word as its proper name.

63. *Hume* (Baumann 1868–9, 2: 565) has already mentioned such a means: "If two numbers are so combined that the one always has a unit which corresponds to each unit of the other, then we claim they are equal." In more recent times, the opinion seems to have found much sympathy among mathematicians, that the equality of numbers must be defined in terms of a one-to-one correspondence. Immediately, however, there arise certain logical hesitations and difficulties, which we must not pass by without examination.

The relationship of equality does not hold only among numbers. It seems to follow from this that the relationship should be defined especially for numbers. One would think it possible to derive a criterion of when numbers are identical with one another from a previously determined concept of identity together with the concept of number, without its being necessary, for this purpose, to define a special concept of numerical identity.

Contrary to this, it should be noted that, for us, the concept of number has not yet been defined, but rather is to be determined by means of our definition of numerical identity. We intend to reconstruct the content of judgments interpretable as expressing identities each side of which is a number. We do not, therefore, want to define equality especially for this instance, but we wish rather, by means of the already familiar concept of equality, to determine that which is to be considered equal. This seems indeed to be a very unusual type of definition, which has probably not yet received sufficient attention from the logicians. Nevertheless, that it is not entirely unheard of may be shown by a few examples:

64. The judgment: 'the [straight] line a is parallel to the [straight] line b', or, symbolically:

$$a \,||\, b,$$

can be interpreted as an equation. If we do this, we obtain the concept of direction and say: 'the direction of line a is the same as the direction of line b'. Hence, we replace the symbol '$||$' by the more general '$=$', by distributing the particular content of the former to a and b. We split up the content in some way other than the original way and thus obtain a new concept. Often the situation is interpreted conversely, and several teachers define: parallel lines are those having the same direction. The theorem "if two straight lines are parallel to a third, then they are parallel to one another" can then be very easily proved on the basis of the similarly worded equality theorem. Unfortunately, this method reverses the natural order of things. For everything geometric must indeed be intuitive, at least originally. Now I ask whether anyone has ever had an intuition of the direction of a straight line? Of the straight line, yes, but can one also distinguish intuitively this line from its direction? Rather difficult! This concept is found only by means of a mental activity connected with intuition. On the other hand, one has a picture of parallel lines. That proof comes about only through a trick in which what is to be proved is covertly presupposed in the use of the word 'direction'; for, were the statement: 'if two straight lines are parallel to a third, then they are parallel to one another' false, then one could not change '$a \| b$' into an equation.

Thus one can obtain from the parallelism of planes a concept which corresponds to that of direction among straight lines. I have seen the name 'orientation' used for this concept. From geometric similarity there arises the concept of shape, so that, e.g., instead of 'the two triangles are similar', one says: 'the two triangles have the same shape' or 'the shape of the one triangle is equal to the shape of the other.' Similarly one can also obtain from the collinear relationship of geometric figures a concept for which a name is probably still lacking.

65. Now, in order to move, e.g., from parallelism[4] to the concept of direction, let us try the following definition: the sentence

'line *a* is parallel to line *b*'

is to be synonymous with

'the direction of line *a* is the same as the direction of line *b*'.

This definition departs from common practice insofar as it apparently defines the already familiar relation of equality, while it should in actuality introduce the expression 'the direction of line *a*', which occurs only incidentally. From this there arises a second hesitation; viz., whether, through such a stipulation, we could not become involved in contradictions with the familiar laws of equality. What are these? They will be developed as analytic truths from the concept itself. Now, Leibniz defines:[5]

[4]In order to be able to express myself more comfortably and to be more easily understood, I speak here of parallelism. The essential parts of these discussions are very easily carried over to the case of numerical equality.

[5]*Non inelegans specimen demonstrandi in abstractis* (Erdmann 1840: 94).

"Eadem sunt, quorum unum potest substitui alteri salva veritate." ["Things are equal which may be substituted for one another without change of truth [value]."]

I will adopt this definition. Whether, like Leibniz, one says 'the same' or 'equal', is of little import. 'The same' does seem to express complete agreement, 'equal' only agreement in this respect or that. One can, however, assume a manner of speaking in which this difference is eliminated, e.g., by saying instead of 'the lines are equal in length' that 'the length of the lines is equal' or 'the same'; instead of saying 'the surfaces are equal in color' one might say 'the color of the surfaces is equal [identical]'.

And this is the way we used the word in the foregoing examples. In fact, all the laws of equality are contained in the principle of universal substitutivity.

In order to justify our proposed definition of the direction of a straight line, we would have to show, then, that

'the direction of *a*'

can be everywhere replaced by

'the direction of *b*',

if line *a* is parallel to line *b*. This is simplified by the fact that, at first, we know no assertion about the direction of a straight line other than its agreement with the direction of another straight line. We would therefore need to demonstrate only the substitutivity in such an equation or in contexts which would contain such equations as component parts.[6] All other statements about directions would have to be defined first, and for these definitions we can adopt the rule that the substitutivity of the direction of a straight line for that of one parallel to it must be preserved.

66. Still a third hesitation arises, however, concerning our proposed definition. In the sentence

'the direction of *a* is equal to the direction of *b*',

the direction of *a* appears as an object,[7] and we have in our definition a means of recognizing this object, should it appear in some other guise,

[6]For example, in a hypothetical judgment an equality of directions could occur either as antecedent or as consequent.

[7]The definite article points to this. A concept is for me a possible predicate in a singular thought content, an object a possible subject of the latter. [Although the terminology of "thought contents" has been adopted, Frege must not be taken to mean anything psychological by 'thought'. For Frege a "thought content" is what is asserted in a statement, asked in a question, etc....] If, in the sentence 'the direction of the axis of the telescope is equal to the direction of the earth's axis', we consider the direction of the telescope's axis to

such as the direction of b. However, this method is not sufficient for all cases. One cannot use it to decide whether England is the same as the direction of the earth's axis. Please excuse this apparently nonsensical example! Naturally, no one is going to confuse England with the direction of the earth's axis; but this is not owing to our definition. The latter says nothing about whether the statement

'the direction of a is equal to q'

is to be affirmed or denied, if q itself is not given in the form 'the direction of b'. We lack the concept of direction; for, if we had this, then we could stipulate that, if q is not a direction, then our statement is to be denied; if q is a direction, then the earlier definition decides. It is now but a step away to define:

q is a direction if there is a straight line b whose direction is q.

However, it is clear that we have now come around in a circle. In order to apply this definition, we would first have to know in each case whether the statement

'q is equal to the direction of b'

was to be affirmed or denied.

67. If we were to say: q is a direction if it is introduced by means of the foregoing definitions, then we would be treating the manner by which the object q is introduced as a property of it, which it is not. The definition of an object, as such, really says nothing about that object; rather it stipulates the meaning of a symbol. Once that has happened, the definition becomes a judgment which treats of the object: it now no longer introduces the object but stands on equal footing with other statements about it. To choose this way out is to presuppose that an object could be given in one way only; otherwise it would not follow from the fact that q is not introduced by means of our definition that it could not be so introduced. The import of any equation would then be that what is given us in the same way should be recognized as the same. But this principle is so obvious and so unfruitful that there is little to be gained by stating it. As a matter of fact, no conclusion could be drawn from it which would not be the same as some premise. The many-sided and broad applicability of equations is based rather on the fact that something is recognizable again even though it is given in a different way.

be the subject, then the predicate is 'equal to the direction of the earth's axis'. This is a concept. But the direction of the earth's axis is only a part of the predicate; the direction is an object, since it can also be made the subject.

68. Since this method fails to yield a sharply delimited concept of direction and, for the same reason, would yield no such concept of number, let us try a different tack. If line a is parallel to line b, then the extension of the concept "line parallel to line a" is the same as the extension of the concept "line parallel to line b"; and conversely: if the extensions of the aforementioned concepts are equal, then a is parallel to b. Let us try, then, to define:

the direction of line a is the extension of the concept "parallel to line a"

the shape of triangle d is the extension of the concept "similar to triangle d."

If we want to apply this to our case, then we must substitute concepts for the lines or the triangles and, for parallelism or similarity, the possibility of correlating in one-to-one fashion the objects falling under the one concept with those falling under the other. As an abbreviation, I will call the concept F equinumerous[8] with the concept G, if this possibility exists; I must, however, request that this word be considered an arbitrarily chosen notational device whose meaning is not to be taken from its linguistic composition, but rather from the foregoing definition.

I define accordingly:

the number which applies to the concept F is the extension[9] of the concept "equinumerous with the concept F."

69. That this definition is correct will, at first perhaps, not be so clear. Don't we mean something other than [different from] a number by the extension of a concept? What we do mean becomes clear from the basic statements that can be made about extensions of concepts. They are the following:

[8][Frege coined 'gleichzählig' for this. In his translation, J. L. Austin (Frege 1950) uses 'equal' and adds the following footnote: "*Gleichzählig* – an invented word, literally 'identi-numerate' or 'tautarithmic'; but these are too clumsy for constant use. Other translators have used 'equinumerous'; 'equinumerate' would be better. Later writers have used 'similar' in this connection (but as a predicate of 'class' not of 'concept')." – Tr.]

[9]I think we could say for 'extension of the concept' simply 'concept'. However, there might be two objections:

1. This stands in contradiction to my earlier assertion that the individual number is an object, the latter being indicated by the use of the article in expressions like "the 2," by the impossibility of speaking about ones, twos, etc. in the plural, and by the fact that the number makes up only a part of the predicate of a statement of number.

2. Concepts can have the same extension without coinciding.

Now I am of the opinion that both these objections can be met, but doing this would lead us too far astray. I presuppose that one knows what the extension of a concept is.

1. that they are equal,
2. that the one encompasses more than the other.

Now the statement

> 'the extension of the concept "equinumerous with the concept F" is the same as the extension of the concept "equinumerous with the concept G"'

is true if and only if the statement

> 'the same number applies to the concept F as to the concept G'

is also true. Hence, there is complete agreement here.

To be sure, one does not say that one number encompasses more than another in the same sense that the extension of one concept encompasses more than does another; however, so is it impossible that

> the extension of the concept "equinumerous with the concept F"

should encompass more than

> the extension of the concept "equinumerous with the concept G"

Rather, if all concepts which are equinumerous with G are also equinumerous with F, then conversely, all concepts which are equinumerous with F are also equinumerous with G. This term 'more encompassing' should not, of course, be confused with the term 'greater', which occurs among numbers.

Certainly, it is also imaginable that the extension of the concept "equinumerous with the concept F" might encompass more or less than the extension of another concept; the latter, then, could not be a number according to our definition. Furthermore, it is not usual to call a number more or less encompassing than the extension of a concept. Nonetheless, there is nothing in the way of so speaking should the occasion arise.

Completion and confirmation of our definition

70. Definitions are confirmed by their fruitfulness. Those definitions which could just as easily be left out without invalidating proofs should be discarded as wholly worthless.

Let us see, then, whether some of the familiar properties of numbers can be derived from our definition of the number which applies to the concept F. We will be satisfied here by the most simple properties.

In order to do this, it is necessary to specify somewhat more exactly the meaning of equinumerosity. We defined it in terms of one-to-one corre-

lation; just how I want to understand this expression must now be explained, since one might easily suspect a connection with intuition.

Let us consider the following example: If a waiter wants to be sure that he is placing just as many knives as plates on the table, he need count neither of them if he places a knife immediately to the right of each plate so that each knife on the table is located to the immediate right of a plate. The plates and knives are thus correlated in one-to-one fashion with one another, in this case through the same positional relationship. If, in the sentence

> 'α lies immediately to the right of A'

we imagine all sorts of objects substituted for α and A, then the part of the content which remains unchanged through all this forms the essence of the relation. Let us generalize this:

When, from a thought content which concerns an object a and an object b, we remove a and b, we retain the concept of a relation, which, accordingly, requires supplementation in two places. If, in the statement

> 'the earth has more mass than the moon',

we remove "the earth," then we obtain the concept "having more mass than the moon." If, on the other hand, we remove the object, "the moon," we gain the concept "having less mass than the earth." Removing both at once leaves a relational concept, which has in itself no more meaning than a simple concept, and which must be supplemented to become a thought content. But this supplementation can come about in various ways: instead of the earth and moon, I can take, e.g., the sun and earth, thus also effecting a removal of the earth and moon [and disclosing the relational nature of the concept].

The individual pairs of associated objects are related – one might say as subjects – to the relational concept in a manner similar to that of the individual object and the concept under which it falls. The subject here is a composite. At times, when the relation is a reversible one [symmetric in two argument places], this is also expressed linguistically, as in the sentence 'Peleus and Thetis were the parents of Achilles'.[10]

On the other hand, it would not be possible to reformulate the statement 'the earth is greater than the moon' so as to make 'the earth and the moon' appear as a compound subject, because the 'and' always indicates a certain equality of rank. This, however, does not affect the matter at hand.

The concept of relation, like the simple concept, belongs, then, to pure

[10]Do not confuse this with the case where the 'and' only seemingly connects the subjects, but in reality, however, connects two sentences.

logic. The particular content of the relation does not concern us here, but only its logical form. And [the truth of] whatever can be asserted about this form is analytic and is known *a priori*. This holds for the relational concepts as well as for the others.

Just as

'a falls under the concept F'

is the general form of a thought content concerning the object a, so can

'a stands in the relation ϕ to b'

be taken as the general form of a thought content concerning objects a and b.

71. Now if each object which falls under the concept F stands in the relation ϕ to an object falling under the concept G, and if, for each object which falls under G, there is an object falling under F which stands in the relation ϕ to it, then the objects falling under F and G are correlated with one another by means of the relation ϕ.

We may still ask what the expression

'each object which falls under F stands in the relation ϕ to an object falling under G'

means, if no object at all falls under F. By this I mean that the two statements

'a falls under F'

and

'a does not stand in the relation ϕ to any object falling under G'

cannot stand together, no matter what a denotes, so that either the first or the second or both are false. From this it follows that if there is no object falling under F, then "each object which falls under F stands in the relation ϕ to an object falling under G," because the first statement

'a falls under F'

is always to be denied, no matter what a might be.

Thus

'for each object which falls under G, there is an object falling under F which stands in the relation ϕ to it'

means that the two statements

'a falls under G'

and

'no object falling under F stands in the relation ϕ to a'

cannot stand together, whatever a may be.

72. We have now seen when the objects falling under the concepts F and G are correlated with one another by means of the relation ϕ. This correlation is here supposed to be one-to-one. By that I mean that the following two statements must hold:

1. If d stands in the relation ϕ to a, and if d stands in the relation ϕ to e, then, no matter what d, a, and e may be, a is always the same as e.
2. If d stands in the relation ϕ to a, and if b stands in the relation ϕ to a, then, whatever d, b, and a may be, d is always the same as b.

By these statements we have reduced one-to-one correlations to purely logical terms and can now offer the following definition:

the expression

'the concept F is equinumerous with the concept G'

is to be synonymous with the expression

'there is a relation ϕ which correlates in one-to-one fashion the objects falling under F with the objects falling under G'.

I [now] repeat [our original definition]:

the number which applies to the concept F is the extension of the concept "equinumerous with the concept F,"

and add to it:

the expression:

'n is a number'

is to be synonymous with the expression

'there is a concept to which the number n applies'.

Thus the concept of number is defined, apparently by means of itself, nevertheless without fallacy, because 'the number which applies to the concept F' has already been defined.

73. We want to show next, then, that the number which applies to the concept F is equal to the number which applies to the concept G, if the concept F is equinumerous with the concept G. This sounds like a tautology, but it is not, since the meaning of the word 'equinumerous' does not follow from its (linguistic) composition, but rather from the foregoing definition.

According to our definition, we must show that the extension of the concept "equinumerous with the concept F" is the same as that of the concept "equinumerous with the concept of G," if the concept F is equinumerous with the concept G. In other words, it must be shown that, under this hypothesis, the following statements always hold:

> 'if the concept H is equinumerous with the concept F, then it is also equinumerous with the concept G';

and

> 'if the concept H is equinumerous with the concept G, then it is also equinumerous with the concept F'.

The upshot of the first statement is that there is a relation which correlates in one-to-one fashion the objects falling under the concept H with those falling under the concept G, if there is a relation ϕ which correlates one-to-one the objects falling under the concept F with those falling under the concept G, and if there is a relation ψ which correlates one-to-one the objects falling under the concept H with those falling under the concept F. The following arrangement of the letters will make this easier to see

$$H\psi F\phi G.$$

Such a relation can in fact be given: it is [that] part of the thought content:

> "there is an object to which c stands in the relation ψ and which stands in the relation ϕ to b"

[which remains] if we remove from it c and b (considering them as the things related). It can be shown that this relation is one-to-one and that it correlates the objects falling under the concept H with those falling under the concept G.

In a similar manner, the other theorem can also be proved.[11] Hopefully, these outlines will suffice to demonstrate that we need not borrow

[11]Similarly for its converse: If the number which applies to the concept F is the same as that which applies to the concept G, then the concept F is equinumerous with the concept G.

here any evidence from intuition, and that something may be done with our definitions.

74. We can now go on to the definitions of the individual numbers.

Because nothing falls under the concept "unequal to itself," I define:

> 0 is the number which applies to the concept "unequal to itself."

Perhaps someone will take exception to my speaking about a concept here. He will perhaps object that a contradiction is contained therein and will recall the old stand-bys, wooden iron and the square circle. To my mind, these are not at all as bad as they are made out to be. Of course, they are not exactly useful, but they can't do any harm, either, as long as one doesn't require that something fall under them; and *that* one does not yet do through the mere usage of the concepts. That a concept contains a contradiction is not always obvious without some examination; but to do that, one must have [the concept] and treat it logically just like any other. All that can be demanded of a concept from the point of view of logic and for rigor in proof procedure is its precise delineation; that, for each object, it be determined whether or not it falls under the concept. This requirement is fully satisfied, then, by concepts containing a contradiction, such as "unequal to itself," for it is known of every object that it does not fall under such a concept.[12]

I use the word 'concept' in such a way that

> 'a falls under the concept F'

is the general form of a thought content, which concerns an object a and which remains decidable, whatever one may put for a. And in this sense,

> 'a falls under the concept "unequal to itself"'

is synonymous with

> 'a is unequal to itself'

or

> 'a is unequal to a'.

In defining 0, I could have taken any other concept under which nothing

[12]Completely different from this is the definition of an object in terms of a concept under which it falls. The expression 'the greatest proper fraction' has, for example, no content, because the definite article carries with it the requirement that it refer to a definite object. On the other hand, the concept, "fraction which is less than 1 and has the property that no fraction which is less than 1 exceeds it in magnitude," is wholly unobjectionable. In fact, in order to prove that there is no such fraction, one even needs this concept, even though it contains a contradiction.

falls. It was up to me, however, to choose one of which this could be purely logically proved, and for this purpose "unequal to itself" presented itself most comfortably, whereby I let the previously presented definition of Leibniz hold, which is also purely logical.

75. We must now be able to prove, by means of what has already been said, that every concept under which nothing falls is equinumerous with any other concept under which nothing falls, and only with such a concept; from which it follows that 0 is the number which applies to such a concept and that no object falls under a concept if the number which applies to that concept is 0.

If we assume that no object falls either under the concept F or under the concept G, then, in order to prove that they are equinumerous, we need a relation ϕ about which the following statements hold:

> 'each object which falls under F stands in the relation ϕ to an object which falls under G; for each object which falls under G there is one falling under F which stands in the relation ϕ to it'.

According to what was said earlier about the meaning of these expressions, every relation fulfills these conditions under our hypotheses; hence also equality, which is, moreover, one-to-one. For, both the foregoing statements required of it hold.

If, on the other hand, an object falls under G, e.g., a, whereas none falls under F, then the two statements

> 'a falls under G'

and

> 'no object falling under F stands in the relation ϕ to a'

hold for every relation ϕ; for, the first holds true according to the first assumption, and the second, according to the second. That is, if there is no object falling under F, then there is also none which would stand in any sort of relation to a. There is, therefore, no relation which would, according to our definition, correlate the objects falling under F with those falling under G; accordingly, the concepts F and G are not equinumerous.

76. I want now to define the relation in which any two adjoining members of the series of natural numbers stand to one another. The statement

> 'there is a concept F and an object x falling under it such that the number which applies to the concept F is n, and that the number

> which applies to the concept "falling under F but not identical with x" is m',

is to be synonymous with

> 'n immediately follows m in the series of natural numbers'.

I am avoiding the expression 'n is *the* number immediately following m', because two theorems would first have to be proved in order to justify the use of the definite article.[13] For the same reason, I am not yet saying here '$n=m+1$'; for, by means of the equals sign, $(m+1)$ is also designated as an object.

77. Now in order to arrive at the number 1, we must first show, that there is something which immediately follows 0 in the series of natural numbers.

Let us consider the concept – or, if you prefer – the predicate 'equal to 0'. 0 falls under this. On the other hand, no object falls under the concept "equal to 0 but not equal to 0," so that 0 is the number which applies to this concept. We have therefore, a concept "equal to 0" and an object 0 falling under it, for which it holds that:

> the number which applies to the concept "equal to 0" is equal to the number which applies to the concept "equal to 0";
>
> the number which applies to the concept "equal to 0 but not equal to 0" is 0.

Therefore, according to our definition, the number which applies to the concept "equal to 0" follows immediately after 0 in the series of natural numbers.

If we define, then,

> 1 is the number which applies to the concept "equal to 0,"

then we can express the last statement so:

> 1 immediately follows 0 in the series of natural numbers.

Perhaps it is not superfluous to note that the definition of 1 does not presuppose any observed fact[14] for its objective legitimacy, for one can easily be confused by the fact that certain subjective conditions must be fulfilled in order to enable us to give the definition, and that sense impressions cause us to do so (cf. Erdmann 1877: 164). This can, nevertheless, be the case without the derived theorems ceasing to be *a priori*. To such conditions belongs the requirement, for example, that blood

[13] See footnote 12.
[14] A proposition that is not general.

flow through the brain in sufficient quantity and of the right concentration – at least as far as we know; however, the truth of our last proposition is independent of that; it continues to hold even if this flow no longer takes place. And even if all reasonable creatures should at some time simultaneously slip into hibernation, the truth of the statement would not, as it were, be suspended for the duration of this sleep, but would remain undisturbed. The truth of a statement is not its being thought.

78. I list here several theorems to be proved by means of our definitions. The reader will easily see how this may be done.

I. If a immediately follows 0 in the series of natural numbers, then $a=1$.

II. If 1 is the number which applies to a concept, then there is an object which falls under that concept.

III. If 1 is the number which applies to a concept F; if the object x falls under the concept F, and if y falls under the concept F, then $x=y$; i.e., x is the same as y.

IV. If an object falls under a concept F and if, from the fact that x falls under the concept F and that y falls under the concept F, it may always be inferred that $x=y$, then 1 is the number which applies to the concept F.

V. The relation that m bears to n, if and only if

"n immediately follows m in the series of natural numbers",

is a one-one relation.

Thus far it has not yet been said that for every number there is another which immediately follows it or is immediately followed by it in the series of natural numbers.

VI. Every number except 0 immediately follows another number in the series of natural numbers.

79. Now in order to be able to prove that every number (n) in the series of natural numbers is immediately followed by a number, one must come up with a concept to which this latter number applies. We choose for this:

"belonging to the series of natural numbers ending with n,"

but we must first define it.

To begin with I shall repeat, in somewhat different words, the definition I gave in my *Begriffsschrift* of following in a series:

The statement

'if every object to which x stands in the relation ϕ falls under the concept F, and if, from the fact that d falls under the concept F, it always follows, no matter what d may be, that every object to which d stands in the relation ϕ falls under the concept F, then y falls under the concept F, no matter what concept F might be',

is to be synonymous with

'y follows x in the ϕ-series'

and with

'x precedes y in the ϕ-series'.

80. Several remarks concerning this definition will not be superfluous here. Since the relation ϕ is left indeterminate, the series is not necessarily to be thought of in the form of a spatial or temporal arrangement, although these cases are not excluded.

Some other definition might be considered more natural, e.g., if, in proceeding from x, we always turn our attention from one object to another, to which it stands in the relation ϕ, and if, in this way, we can finally reach y, then we say that y follows x in the ϕ-series.

This is a way of looking at the matter, not a definition. Whether we reach y in the wanderings of our attention can depend on many subjective incidental circumstances; e.g., on the time we have available or on our knowledge of the things. Whether y follows x in the ϕ-series has, in general, nothing at all to do with our attention and the conditions of its progress, but rather it is a matter of objective fact: just as a green leaf reflects certain light rays whether or not they should meet my eye and summon up a sensation; just as a grain of salt is soluble in water whether or not I put it in water and observe the process; and just as it remains soluble even if it is not possible for me to experiment on it.

By means of my definition, the matter is elevated from the realm of the subjectively possible to that of the objectively definite. Indeed, the fact that from certain statements another statement follows is something objective, something independent of whatever laws may govern the wanderings of our attention; and it makes no difference whether we really make the inference or not. Here we have a criterion which decides the question, wherever it can be asked, even though we might be hindered by external difficulties from judging in individual cases whether it is applicable. That makes no difference to the issue itself.

We need not always run through all the intermediate members, from the initial member up to an object, in order to be sure that the latter follows the former. If, e.g., it is given that, in the ϕ-series, b follows a and c follows b, then we can conclude on the basis of our definition that c follows a, without even knowing the intermediate members.

Only by means of this definition of following in a series does it become possible to reduce the rule of inference from n to $(n+1)$, which apparently is peculiar to mathematics, to general logical laws.

81. Now if we have as our relation ϕ the one in which m is related to n by the statement

> 'n immediately follows m in the series of natural numbers',

then we say instead of 'ϕ-series', 'series of natural numbers'.

I define further:

the statement

> 'y follows x in the ϕ-series or y is the same as x',

is to be synonymous with

> 'y belongs to the ϕ-series starting with x'

and with

> 'x belongs to the ϕ-series ending with y'.

According to this, a belongs to the series of natural numbers ending with n if n either follows a in the series of natural numbers or is equal to a.[15]

82. We must now show that, under a condition still to be stated, the number which applies to the concept

> "belonging to the series of natural numbers ending with n"

immediately follows n in the series of natural numbers. Having this result, we will have proved that there is a number which immediately follows n in the series of natural numbers; i.e., that there is no last member of this series. Obviously, this statement cannot be established empirically or by means of induction.

It would take us too far afield to give the proof itself. We can only give a brief sketch of it here. We must prove:

1. If a immediately follows d in the series of natural numbers, and if the number which applies to the concept

   > "belonging to the series of natural numbers ending with d"

[15] If n is not a number, then only n itself belongs to the series of natural numbers ending with n. One should not object to this expression.

   immediately follows d in the series of natural numbers, then the number which applies to the concept

   > "belonging to the series of natural numbers ending with a"

   immediately follows a in the series of natural numbers.

2. We must prove that what has been asserted about d and a in the foregoing statements holds for 0, and then show that it also holds for n, if n belongs to the series of natural numbers beginning with 0. This will result from an application of my definition of

   > 'y follows x in the series of natural numbers',

   taking as the concept F the relation asserted above to hold between d and a, and substituting 0 and n for d and a.

83. In order to prove Theorem 1 of the last paragraph, we must show that a is the number which applies to the concept "belonging to the series of natural numbers ending with a, but not equal to a." And to this end, we must prove that this concept has the same extension as the concept "belonging to the series of natural numbers ending with d." For this, we need the theorem that no object which belongs to the series of natural numbers beginning with 0 can follow itself in the series of natural numbers. The latter must likewise be proved by means of our definition of following in a series, as it is outlined above.[16]

For this reason, we must add the condition that n belong to the series of natural numbers beginning with 0 to the statement that the number which applies to the concept

> "belonging to the series of natural numbers ending with n,"

immediately follows n in the series of natural numbers. There is a shorter way of putting this, which I shall now define:

the statement

> 'n belongs to the series of natural numbers beginning with 0'

is to be synonymous with

> 'n is a finite number'.

We can now express the last theorem thus: no finite number follows itself in the series of natural numbers.

[16] E. Schröder (1873: 63) seems to look upon this theorem as the consequence of an ambiguous terminology. The difficulty which infects his whole presentation of the matter emerges here too; i.e., it is never quite clear whether the number is a symbol and, if so, what its meaning is, or whether it *is* this very meaning. From the fact that one sets up different symbols, so that the same one never recurs, it does not follow that these symbols mean different things.

Infinite numbers

84. In contrast to the finite numbers there are the infinite ones. The number which applies to the concept "finite number" is an infinite one. Let us denote it, say, by $\aleph_0$.[17] Were it a finite number, it could not follow itself in the series of natural numbers. One can show, however, that $\aleph_0$ does just this.

There is nothing somehow mysterious or marvellous about the infinite number $\aleph_0$ when so defined. 'The number which applies to the concept *F* is $\aleph_0$' says nothing more nor less than: there is a relation which establishes a one-to-one correlation between the objects falling under the concept *F* and the finite numbers. This has, according to our definitions, a completely clear and unambiguous sense, and that suffices to justify the use of the symbol $\aleph_0$ and to guarantee it a meaning. That we can form no mental picture of an infinite number is wholly irrelevant and would hold true of finite numbers as well. In this way, our number $\aleph_0$ is something just as determinate as any finite number: it can be recognized without a doubt as the same and differentiated from any other.

85. Recently, in a noteworthy paper (1883b), G. Cantor introduced infinite numbers. I agree with him completely in his evaluation of the view which would have only the finite numbers qualify as real. Neither these nor the fractions are sensibly perceptible and spatial, nor are the negative, irrational, and complex numbers. And if one calls real [only] that which affects the senses, or at least can have sense impressions as an immediate or distant consequence, then certainly none of these numbers is real. But we don't need such sense impressions as evidence for our theorems. A name or a symbol, which is introduced in a logically unobjectionable way, may be used by us without hesitation in our investigations, and thus our number $\aleph_0$ is just as firmly grounded as 2 or 3.

Although I believe I agree with Cantor in this matter, I do, however, deviate from him in terminology. He calls my numbers 'powers', whereas his concept[18] of number is based on ordering. To be sure, finite numbers end up being independent of order; however, this does not hold for infinite numbers. Now the linguistic usage of the word 'number' and of the question 'how many?' contains no indication of a definite order. Cantor's number answers rather the question: 'the last member is the how-manyth member of the sequence?' Therefore my terminology seems to me to

[17] [Frege used '∞_1', but we adopt the aleph notation as being more in keeping with current practice. - Tr.]

[18] This expression may appear to contradict [my earlier remarks emphasizing] the objectivity of concepts; however, only the *terminology* is subjective here.

agree better with linguistic usage. If one extends the meaning of a word, then one must take care that as many general statements as possible retain their validity, and particularly statements as basic as, for instance, [the one asserting] for numbers their independence of the sequence. We have needed no extension at all, because our concept of number immediately embraces infinite numbers as well.

86. In order to obtain his infinite numbers, Cantor introduces the relational concept of following in a sequence, which differs from my "following in a series." According to him, for instance, a sequence would result if one were so to order the finite positive whole numbers that the odd numbers followed one another in their own natural order, and similarly the even numbers in theirs, and it were further stipulated that all the even numbers should come after all the odd numbers. In this sequence, e.g., 0 would follow 13. There would, however, be no number immediately preceding 0. Now this case cannot occur within my definition of following in a series. It may be strictly proved, without using intuition, that, if y follows x in the ϕ-series, there is an object which immediately precedes y in this series. It seems to me, then, that exact definitions of following in a sequence and of number [in Cantor's sense] are still lacking. Thus Cantor bases himself on a somewhat mysterious "inner intuition" where a proof from definitions should be striven for and would probably be found. For I think I can foresee how those concepts could be defined. In any case, I in no way wish these comments to be taken as an attack on the justifiability or fruitfulness of these concepts. On the contrary, I welcome these investigations as an extension of the science, especially because they strike a purely arithmetic path to higher infinite numbers (powers).

Conclusion

87. I hope in this monograph to have made it probable that arithmetic laws are analytic judgments, and therefore *a priori.* According to this, arithmetic would be only a further developed logic, every arithmetic theorem a logical law, albeit a derived one. The applications of arithmetic to the explanation of natural phenomena would be logical processing of observed facts;[19] computation would be inference. Numerical laws will not need, as Baumann (1868–9, 2: 670) contends, a practical confirmation in order to be applicable in the external world; for, in the external world, the totality of space and its contents, there are no concepts, no properties of concepts, no numbers. Therefore, the numerical laws are

[19] Observation itself already includes a logical activity.

really not applicable to the external world: they are not laws of nature. They are, however, applicable to judgments, which are true of things in the external world: they are laws of the laws of nature. They assert connections not between natural phenomena, but rather between judgments; and it is to the latter that the laws of nature belong.

88. Kant (1867–8, 3: 39ff) evidently underestimated the value of analytic judgments – probably as the result of having too narrow a definition of the concept – although he apparently also had in mind the broader concept used here.[20] Taking his definition as a basis, the division of judgments into the analytic and the synthetic is not exhaustive. He is thinking of universal affirmative judgments. In such cases, one can speak of a concept of the subject and inquire whether the concept of the predicate – as would result from *his* definition – is contained in it. How can we do this, however, when the subject is a single object? Or when the judgment is existential? In such cases there can be, in Kant's sense, no talk of a concept of the subject. Kant seems to have thought of the concept as determined by subordinate characteristics; that, however, is one of the least fruitful notions of concept. If one surveys the foregoing definitions, one will hardly find one of this kind. The same is true of the really fruitful definitions in mathematics, e.g., of the continuity of a function. There we don't have a series of subordinate characteristics but rather a more intimate, I should say more organic, connection between the [elements of the] definitions. The difference can be illustrated by means of a geometrical analogy. If the concepts (or their extensions) are represented by regions of a plane, then the concept defined by means of subordinate characteristics corresponds to the region which is the overlap of all the individual regions corresponding to these characteristics; it is enclosed by parts of their boundaries. Pictorially speaking, in such a definition, we delimit a region by using in a new way lines already given. In doing this, however, nothing essentially new comes out. The more fruitful definitions draw border lines which had not previously been given. What can be inferred from them cannot be seen in advance; one does not simply withdraw again from the box what one has put into it. These inferences expand our knowledge and one should, therefore, following Kant, consider them synthetic. Nevertheless, they can be proved purely logically and hence are analytic. They are in fact contained in the definitions, but like the plant in the seed, not like the rafter in the house. One often needs several definitions to prove a theorem, which consequently is contained in no single definition, but nevertheless follows in a purely logical way from all of them together.

[20][Kant] says that a synthetic statement can be understood according to the Theorem of Contradiction only if another synthetic statement is presupposed (1867–8, 3: 43).

89. I must also contradict the generality of Kant's assertion (1867–8, 3: 82) that without sensible perception no object would be given us. Zero and 1 are objects that cannot be given us sensibly. And those who hold the smaller numbers to be intuitive will surely have to concede that none of the numbers greater than $1000^{1000^{1000}}$ can be given them intuitively, and that we nevertheless know a good deal about them. Perhaps Kant was using the word 'object' in a somewhat different sense; but then zero, 1, and our $\aleph_0$ disappear completely from his considerations; for, they are not concepts either, and Kant demands even of concepts that their objects be appended to them in intuition.

In order not to open myself to the criticism of carrying on a picayune search for faults in the work of a genius whom we look up to only with thankful awe, I believe I should also emphasize our areas of agreement, which are far more extensive than those of our disagreement. To touch on only the immediate points, I see a great service in Kant's having distinguished between synthetic and analytic judgments. In terming geometric truths synthetic and *a priori,* he uncovered their true essence. And this is still worth repeating today, because it is still often not recognized. If Kant erred with respect to arithmetic, this does not detract essentially, I think, from his merit. It was important for him that there should be synthetic judgments *a priori;* whether they occur only in geometry or also in arithmetic is of little importance.

90. I do not claim to have made the analytic nature of arithmetic theorems more than probable, because one can always still doubt whether their proof can be carried out completely from purely logical laws, whether evidence of another sort has not crept in unnoticed somewhere. This doubt is also not entirely relieved by the outlines which I have given of the proofs of a few theorems; it can only be alleviated by an airtight chain of reasoning, such that no step is made which is not in conformity with one of a few rules of inference recognized as purely logical. Thus until now, hardly a single [real] proof has ever been offered, because the mathematician is satisfied if every transition to a new judgment appears to him to be correct, without asking whether this appearance is logical or intuitive. A step in such a proof is often quite complex and involves several simple inferences, in addition to which intuitive considerations can creep in. One proceeds in jumps, and from this there arises the impression of an over-rich variety of rules of inference used in mathematics. For, the greater the jumps, the more complex are the combinations of simple inferences and intuitive axioms which they can represent. Nevertheless, such a transition often occurs to us directly, without our being conscious of the intermediate steps, and since it does not present itself as one of the recognized logical rules of inference, we are immediately ready to consider this manifest transition as an intuitive one and the inferred truth as a synthetic one, even when the range of its validity extends far beyond intuition.

Proceeding in this way, it is not possible clearly to separate the synthetic, based on intuition, from the analytic. Nor will it be possible to compile with completeness and certainty the axioms of intuition needed to make every mathematical proof capable of proceeding from these axioms alone, according to logical laws.

91. The requirement of avoiding all jumps in a proof must, therefore, be imposed. That it is so difficult to satisfy is owing to the tediousness of a step-by-step procedure. Every proof, which is even slightly involved, threatens to become enormously long. In addition to this, the superfluity of logical forms expressed in language makes it difficult to extract a group of rules of inference sufficient for all cases and yet easy to survey.

In order to minimize the effects of these drawbacks, I have devised my concept writing. It strives for greater brevity and comprehensibility of expression and is manipulated in a few standard ways, as in a computation, so that no transition is permitted which does not conform to rules set up once for all.[21] No assumption can then slip in unnoticed. I have thus proved a theorem,[22] borrowing no axioms from intuition, which one would consider at first glance to be synthetic and which I shall state here as follows:

If the relation of each member of a series to its successor is one-to-one, and if *m* and *y* follow *x* in this series, then *y* precedes *m* in this series, or coincides with it, or follows *m*.

From this proof, one can see that theorems which expand our knowledge can contain analytic judgments.[23]

[21] It is, however, supposed to be able to express not only the logical form of a statement, as does the Boolean notation, but also its content.

[22] *Begriffsschrift*, 1879, p. 86, Formula 133.

[23] This proof will be found to be still much too lengthy, a disadvantage which may seem to more than balance out the almost unconditional guarantee against a mistake or a loophole. My purpose at that time was to reduce everything to the smallest possible number of the simplest possible logical laws. As a result of this, I applied only one rule of inference. I pointed out even then, in the foreword (p. vii) that, for further application, it would be recommended to admit more rules of inference. This can be done without impairing the validity of the chain of reasoning, and an important abbreviation could thereby be achieved.

[Recapitulation]

106. Let us cast a quick glance backward on the course of our investigation. After determining that a number was not a collection of things nor a property of such a collection, nor, furthermore, the subjective product of mental processes, we decided that a statement of number asserts something objective about a concept. We defined first the individual numbers 0, 1 etc., and then following in the number series. Our first attempt failed, because in it we stated the meaning of only whole assertions about concepts, and not of 0 and 1 separately, although these entered into those assertions. As a result of this, we could not prove the equality of numbers. It was shown that the numbers with which arithmetic concerns itself must be understood not as dependent attributes, but rather substantivally.[24] Thus numbers appeared to us as recognizable objects, although not physical ones nor even merely spatial ones, nor ones which we could picture in imagination. We then established the basic theorem: that the meaning of a word is not to be defined separately, but rather in the context of a statement; only by following this theorem can we, I think, avoid the physical interpretation of number, without slipping into psychological interpretation. There is only one type of statement which must have a sense for every object; that is the recognition sentences, called equations in the case of numbers. We saw that statements of number are also to be interpreted as equations. It became a question, then, of determining the sense of a numerical equation and of expressing this sense without making use of the number-words or the word 'number'. The possibility of establishing a one-to-one correspondence between the objects falling under a concept *F* and those falling under a concept *G* was found to be the content of a recognition judgment about numbers. Our definition, therefore, had to posit that possibility as synonymous with a numerical equation. We recalled similar instances: the definition of direction from parallelism, of shape from similarity, etc.

107. The question then arose: when are we justified in interpreting a content to be that of a recognition judgment? For this, the condition must be fulfilled that in every judgment the left side of the tentatively assumed equation can be replaced by the right, without altering the truth of the judgment. Now, at first and without resorting to further definitions, no further assertion about the left or right side of such an equation is known to us beyond the assertion of their equality. Substitutivity had therefore to be proved only for equations.

[24] The difference corresponds to that between 'blue' and 'the color of the sky'.

A doubt still remained, however. A recognition statement must always have a sense. If we interpreted the possibility of correlating in one-to-one fashion the objects falling under the concept F with those falling under the concept G as an equation, by saying for it: 'the number which applies to the concept F is equal to the number which applies to the concept G' and thereby introducing the expression 'the number which applies is the concept F', then we have a sense for the equation only if both sides have the form just mentioned. We would not be able to judge according to such a definition whether an equation only one side of which had this form was true or false. That caused us to make the following definition:

> The number which applies to the concept F is the extension of the concept "concept equinumerous with the concept F,"

by which we called a concept F equinumerous with a concept G, if there exists the possibility of correlating them one-to-one.

In doing this, we presuppose that the sense of the expression 'extension of a concept' is familiar. This method of overcoming the difficulty will probably not be everywhere applauded, and some will prefer to set aside this doubt in another way. I, too, place no decisive weight on the introduction of the extension of a concept.

108. We still had to define one-to-one correspondences; we reduced them to purely logical terms. After we had outlined the proof of the theorem that the number which applies to the concept F is equal to that which applies to the concept G, if the concept F is equinumerous with the concept G, we defined 0, the expression 'n immediately follows m in the series of natural numbers', and the number 1, and we showed that 1 immediately follows 0 in the series of natural numbers. We presented a few theorems which could be easily proved at this point and then went somewhat more deeply into the following, which demonstrates the infinity of the number series:

> Every number in the series of natural numbers is followed by a number.

We were thereby led to the concept "belonging to the series of natural numbers ending with n," of which we wanted to show that the number applying to it immediately follows n in the series of natural numbers. We defined it at first by means of an object y following an object x in a general ϕ-series. The sense of this expression was also reduced to purely logical terms. And thereby we succeeded in proving that the rule of inference from n to $(n+1)$, which is usually considered a peculiarly mathematical one, is based on the general logical rules of inference.

For the proof of the infinity of the number series, we needed the theorem that no finite number follows itself in the series of natural numbers. We thus arrived at the concepts of finite and infinite numbers. We showed that the latter is basically no less justified logically than is the former. For the purposes of comparison, we drew upon Cantor's infinite numbers and his "following in a sequence," where we pointed out the difference in terminology.

109. We thus rendered the analytic and *a priori* character of arithmetic truths highly probable, arriving at an improvement on Kant's point of view. We saw further what was still lacking in order to elevate that probability to certainty and we indicated the path that must lead to this.

Selections from *Introduction to Mathematical Philosophy*

BERTRAND RUSSELL

I. The series of natural numbers

Mathematics is a study which, when we start from its most familiar portions, may be pursued in either of two opposite directions. The more familiar direction is constructive, towards gradually increasing complexity: from integers to fractions, real numbers, complex numbers; from addition and multiplication to differentiation and integration, and on to higher mathematics. The other direction, which is less familiar, proceeds, by analysing, to greater and greater abstractness and logical simplicity; instead of asking what can be defined and deduced from what is assumed to begin with, we ask instead what more general ideas and principles can be found, in terms of which what was our starting-point can be defined or deduced. It is the fact of pursuing this opposite direction that characterises mathematical philosophy as opposed to ordinary mathematics. But it should be understood that the distinction is one, not in the subject matter, but in the state of mind of the investigator. Early Greek geometers, passing from the empirical rules of Egyptian land-surveying to the general propositions by which those rules were found to be justifiable, and thence to Euclid's axioms and postulates, were engaged in mathematical philosophy, according to the above definition; but when once the axioms and postulates had been reached, their deductive employment, as we find it in Euclid, belonged to mathematics in the ordinary sense. The distinction between mathematics and mathematical philosophy is one which depends upon the interest inspiring the research, and upon the stage which the research has reached; not upon the propositions with which the research is concerned.

We may state the same distinction in another way. The most obvious and easy things in mathematics are not those that come logically at the beginning; they are things that, from the point of view of logical deduction, come somewhere in the middle. Just as the easiest bodies to see are those that are neither very near nor very far, neither very small nor very

Reprinted by kind permission of the publishers from Bertrand Russell, *Introduction to Mathematical Philosophy* (New York: The Macmillan Company; London: George Allen & Unwin Ltd., 1919), pp. 1–19, 194–206.

great, so the easiest conceptions to grasp are those that are neither very complex nor very simple (using "simple" in a *logical* sense). And as we need two sorts of instruments, the telescope and the microscope, for the enlargement of our visual powers, so we need two sorts of instruments for the enlargement of our logical powers, one to take us forward to the higher mathematics, the other to take us backward to the logical foundations of the things that we are inclined to take for granted in mathematics. We shall find that by analysing our ordinary mathematical notions we acquire fresh insight, new powers, and the means of reaching whole new mathematical subjects by adopting fresh lines of advance after our backward journey. It is the purpose of this book to explain mathematical philosophy simply and untechnically, without enlarging upon those portions which are so doubtful or difficult that an elementary treatment is scarcely possible. A full treatment will be found in *Principia Mathematica* (1910–13); the treatment in the present volume is intended as an introduction.

To the average educated person of the present day, the obvious starting-point of mathematics would be the series of whole numbers,

$$1, 2, 3, 4, \ldots, \text{etc.}$$

Probably only a person with some mathematical knowledge would think of beginning with 0 instead of with 1, but we will presume this degree of knowledge; we will take as our starting-point the series:

$$0, 1, 2, 3, \ldots n, n+1, \ldots$$

and it is this series that we shall mean when we speak of the "series of natural numbers."

It is only at a high stage of civilisation that we could take this series as our starting-point. It must have required many ages to discover that a brace of pheasants and a couple of days were both instances of the number 2: the degree of abstraction involved is far from easy. And the discovery that 1 is a number must have been difficult. As for 0, it is a very recent addition; the Greeks and Romans had no such digit. If we had been embarking upon mathematical philosophy in earlier days, we should have had to start with something less abstract than the series of natural numbers, which we should reach as a stage on our backward journey. When the logical foundations of mathematics have grown more familiar, we shall be able to start further back, at what is now a late stage in our analysis. But for the moment the natural numbers seem to represent what is easiest and most familiar in mathematics.

But though familiar, they are not understood. Very few people are prepared with a definition of what is meant by "number," or "0," or "1."

It is not very difficult to see that, starting from 0, any other of the natural numbers can be reached by repeated additions of 1, but we shall have to define what we mean by "adding 1," and what we mean by "repeated." These questions are by no means easy. It was believed until recently that some, at least, of these first notions of arithmetic must be accepted as too simple and primitive to be defined. Since all terms that are defined are defined by means of other terms, it is clear that human knowledge must always be content to accept some terms as intelligible without definition, in order to have a starting-point for its definitions. It is not clear that there must be terms which are *incapable* of definition: it is possible that, however far back we go in defining, we always *might* go further still. On the other hand, it is also possible that, when analysis has been pushed far enough, we can reach terms that really are simple, and therefore logically incapable of the sort of definition that consists in analysing. This is a question which it is not necessary for us to decide; for our purposes it is sufficient to observe that, since human powers are finite, the definitions known to us must always begin somewhere, with terms undefined for the moment, though perhaps not permanently.

All traditional pure mathematics, including analytical geometry, may be regarded as consisting wholly of propositions about the natural numbers. That is to say, the terms which occur can be defined by means of the natural numbers, and the propositions can be deduced from the properties of the natural numbers – with the addition, in each case, of the ideas and propositions of pure logic.

That all traditional pure mathematics can be derived from the natural numbers is a fairly recent discovery, though it had long been suspected. Pythagoras, who believed that not only mathematics, but everything else, could be deduced from numbers, was the discoverer of the most serious obstacle in the way of what is called the "arithmetising" of mathematics. It was Pythagoras who discovered the existence of incommensurables, and, in particular, the incommensurability of the side of a square and the diagonal. If the length of the side is 1 inch, the number of inches in the diagonal is the square root of 2, which appeared not to be a number at all. The problem thus raised was solved only in our own day, and was only solved *completely* by the help of the reduction of arithmetic to logic, which will be explained in following chapters. For the present, we shall take for granted the arithmetisation of mathematics, though this was a feat of the very greatest importance.

Having reduced all traditional pure mathematics to the theory of the natural numbers, the next step in logical analysis was to reduce this theory itself to the smallest set of premisses and undefined terms from which it could be derived. This work was accomplished by Peano. He showed

that the entire theory of the natural numbers could be derived from three primitive ideas and five primitive propositions in addition to those of pure logic. These three ideas and five propositions thus became, as it were, hostages for the whole of traditional pure mathematics. If they could be defined and proved in terms of others, so could all pure mathematics. Their logical "weight," if one may use such an expression, is equal to that of the whole series of sciences that have been deduced from the theory of the natural numbers; the truth of this whole series is assured if the truth of the five primitive propositions is guaranteed, provided, of course, that there is nothing erroneous in the purely logical apparatus which is also involved. The work of analysing mathematics is extraordinarily facilitated by this work of Peano's.

The three primitive ideas in Peano's arithmetic are:

0, number, successor.

By "successor" he means the next number in the natural order. That is to say, the successor of 0 is 1, the successor of 1 is 2, and so on. By "number" he means, in this connection, the class of natural numbers.[1] He is not assuming that we know all the members of this class, but only that we know what we mean when we say that this or that is a number, just as we know what we mean when we say "Jones is a man," though we do not know all men individually.

(1) 0 is a number.
(2) The successor of any number is a number.
(3) No two numbers have the same successor.
(4) 0 is not the successor of any number.
(5) Any property which belongs to 0, and also to the successor of every number which has the property, belongs to all numbers.

The last of these is the principle of mathematical induction. We shall have much to say concerning mathematical induction in the sequel; for the present, we are concerned with it only as it occurs in Peano's analysis of arithmetic.

Let us consider briefly the kind of way in which the theory of the natural numbers results from these three ideas and five propositions. To begin with, we define 1 as "the successor of 0," 2 as "the successor of 1," and so on. We can obviously go on as long as we like with these definitions, since, in virtue of (2), every number that we reach will have a successor, and, in virtue of (3), this cannot be any of the numbers already defined, because, if it were, two different numbers would have the same

[1] We shall use "number" in this sense in the present chapter. Afterwards the word will be used in a more general sense.

successor; and in virtue of (4) none of the numbers we reach in the series of successors can be 0. Thus the series of successors gives us an endless series of continually new numbers. In virtue of (5) all numbers come in this series, which begins with 0 and travels on through successive successors: for (*a*) 0 belongs to this series, and (*b*) if a number n belongs to it, so does its successor, whence, by mathematical induction, every number belongs to the series.

Suppose we wish to define the sum of two numbers. Taking any number m, we define $m+0$ as m, and $m+(n+1)$ as the successor of $m+n$. In virtue of (5) this gives a definition of the sum of m and n, whatever number n may be. Similarly we can define the product of any two numbers. The reader can easily convince himself that any ordinary elementary proposition of arithmetic can be proved by means of our five premisses, and if he has any difficulty he can find the proof in Peano.

It is time now to turn to the considerations which make it necessary to advance beyond the standpoint of Peano, who represents the last perfection of the "arithmetisation" of mathematics, to that of Frege, who first succeeded in "logicising" mathematics, *i.e.* in reducing to logic the arithmetical notions which his predecessors had shown to be sufficient for mathematics. We shall not, in this chapter, actually give Frege's definition of number and of particular numbers, but we shall give some of the reasons why Peano's treatment is less final than it appears to be.

In the first place, Peano's three primitive ideas – namely, "0," "number," and "successor" – are capable of an infinite number of different interpretations, all of which will satisfy the five primitive propositions. We will give some examples.

(1) Let "0" be taken to mean 100, and let "number" be taken to mean the numbers from 100 onward in the series of natural numbers. Then all our primitive propositions are satisfied, even the fourth, for, though 100 is the successor of 99, 99 is not a "number" in the sense which we are now giving to the word "number." It is obvious that any number may be substituted for 100 in this example.

(2) Let "0" have its usual meaning, but let "number" mean what we usually call "even numbers," and let the "successor" of a number be what results from adding two to it. Then "1" will stand for the number two, "2" will stand for the number four, and so on; the series of "numbers" now will be

0, two, four, six, eight ...

All Peano's five premisses are satisfied still.

(3) Let "0" mean the number one, let "number" mean the set

$$1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots$$

and let "successor" mean "half". Then all Peano's five axioms will be true of this set.

It is clear that such examples might be multiplied indefinitely. In fact, given any series

$$x_0, x_1, x_2, x_3, \ldots x_n, \ldots$$

which is endless, contains no repetitions, has a beginning, and has no terms that cannot be reached from the beginning in a finite number of steps, we have a set of terms verifying Peano's axioms. This is easily seen, though the formal proof is somewhat long. Let "0" mean x_0, let "number" mean the whole set of terms, and let the "successor" of x_n mean x_{n+1}. Then

(1) "0" is a number," i.e. x_0 is a member of the set.
(2) "The successor of any number is a number," i.e. taking any term x_n in the set, x_{n+1} is also in the set.
(3) "No two numbers have the same successor," i.e. if x_m and x_n are two different members of the set, x_{m+1} and x_{n+1} are different; this results from the fact that (by hypothesis) there are no repetitions in the set.
(4) "0 is not the successor of any number," i.e. no term in the set comes before x_0.
(5) This becomes: Any property which belongs to x_0, and belongs to x_{n+1} provided it belongs to x_n, belongs to all the x's.

This follows from the corresponding property for numbers.

A series of the form

$$x_0, x_1, x_2, \ldots x_n, \ldots$$

in which there is a first term, a successor to each term (so that there is no last term), no repetitions, and every term can be reached from the start in a finite number of steps, is called a *progression*. Progressions are of great importance in the principles of mathematics. As we have just seen, every progression verifies Peano's five axioms. It can be proved, conversely, that every series which verifies Peano's five axioms is a progression. Hence these five axioms may be used to define the class of progressions: "progressions" are "those series which verify these five axioms." Any progression may be taken as the basis of pure mathematics: we may give the name "0" to its first term, the name "number" to the whole set of its terms, and the name "successor" to the next in the progression. The progression need not be composed of numbers: it may be composed of points in space, or moments of time, or any other terms of which there is an infinite supply. Each different progression will give rise to a different interpretation of all the propositions of traditional pure mathematics; all these possible interpretations will be equally true.

In Peano's system there is nothing to enable us to distinguish between these different interpretations of his primitive ideas. It is assumed that we know what is meant by "0," and that we shall not suppose that this symbol means 100 or Cleopatra's Needle or any of the other things that it might mean.

This point, that "0" and "number" and "successor" cannot be defined by means of Peano's five axioms, but must be independently understood, is important. We want our numbers not merely to verify mathematical formulae, but to apply in the right way to common objects. We want to have ten fingers and two eyes and one nose. A system in which "1" meant 100, and "2" meant 101, and so on, might be all right for pure mathematics, but would not suit daily life. We want "0" and "number" and "successor" to have meanings which will give us the right allowance of fingers and eyes and noses. We have already some knowledge (though not sufficiently articulate or analytic) of what we mean by "1" and "2" and so on, and our use of numbers in arithmetic must conform to this knowledge. We cannot secure that this shall be the case by Peano's method; all that we can do, if we adopt his method, is to say "we know what we mean by '0' and 'number' and 'successor,' though we cannot explain what we mean in terms of other simpler concepts." It is quite legitimate to say this when we must, and at *some* point we all must; but it is the object of mathematical philosophy to put off saying it as long as possible. By the logical theory of arithmetic we are able to put it off for a very long time.

It might be suggested that, instead of setting up "0" and "number" and "successor" as terms of which we know the meaning although we cannot define them, we might let them stand for *any* three terms that verify Peano's five axioms. They will then no longer be terms which have a meaning that is definite though undefined: they will be "variables," terms concerning which we make certain hypotheses, namely, those stated in the five axioms, but which are otherwise undetermined. If we adopt this plan, our theorems will not be proved concerning an ascertained set of terms called "the natural numbers," but concerning all sets of terms having certain properties. Such a procedure is not fallacious; indeed for certain purposes it represents a valuable generalisation. But from two points of view it fails to give an adequate basis for arithmetic. In the first place, it does not enable us to know whether there are any sets of terms verifying Peano's axioms; it does not even give the faintest suggestion of any way of discovering whether there are such sets. In the second place, as already observed, we want our numbers to be such as can be used for counting common objects, and this requires that our numbers should have a *definite* meaning, not merely that they should have certain formal properties. This definite meaning is defined by the logical theory of arithmetic.

II. Definition of number

The question "What is a number?" is one which has often been asked but has only been correctly answered in our own time. The answer was given by Frege in 1884, in his *Grundlagen der Arithmetik.*[2] Although this book is quite short, not difficult, and of the very highest importance, it attracted almost no attention, and the definition of number which it contains remained practically unknown until it was rediscovered by the present author in 1901.

In seeking a definition of number, the first thing to be clear about is what we may call the grammar of our inquiry. Many philosophers, when attempting to define number, are really setting to work to define plurality, which is quite a different thing. *Number* is what is characteristic of numbers, as *man* is what is characteristic of men. A plurality is not an instance of number, but of some particular number. A trio of men, for example, is an instance of the number 3, and the number 3 is an instance of number; but the trio is not an instance of number. This point may seem elementary and scarcely worth mentioning; yet it has proved too subtle for the philosophers, with few exceptions.

A particular number is not identical with any collection of terms having that number: the number 3 is not identical with the trio consisting of Brown, Jones, and Robinson. The number 3 is something which all trios have in common, and which distinguishes them from other collections. A number is something that characterises certain collections, namely, those that have that number.

Instead of speaking of a "collection," we shall as a rule speak of a "class," or sometimes a "set." Other words used in mathematics for the same thing are "aggregate" and "manifold." We shall have much to say later on about classes. For the present, we will say as little as possible. But there are some remarks that must be made immediately.

A class or collection may be defined in two ways that at first sight seem

[2]The same answer is given more fully and with more development in his *Grundgesetze der Arithmetik,* vol. 1, 1893.

quite distinct. We may enumerate its members, as when we say, "The collection I mean is Brown, Jones, and Robinson." Or we may mention a defining property, as when we speak of "mankind" or "the inhabitants of London." The definition which enumerates is called a definition by "extension," and the one which mentions a defining property is called a definition by "intension." Of these two kinds of definition, the one by intension is logically more fundamental. This is shown by two considerations: (1) that the extensional definition can always be reduced to an intensional one; (2) that the intensional one often cannot even theoretically be reduced to the extensional one. Each of these points needs a word of explanation.

(1) Brown, Jones, and Robinson all of them possess a certain property which is possessed by nothing else in the whole universe, namely, the property of being either Brown or Jones or Robinson. This property can be used to give a definition by intension of the class consisting of Brown and Jones and Robinson. Consider such a formula as "x is Brown or x is Jones or x is Robinson." This formula will be true for just three x's, namely, Brown and Jones and Robinson. In this respect it resembles a cubic equation with its three roots. It may be taken as assigning a property common to the members of the class consisting of these three men, and peculiar to them. A similar treatment can obviously be applied to any other class given in extension.

(2) It is obvious that in practice we can often know a great deal about a class without being able to enumerate its members. No one man could actually enumerate all men, or even all the inhabitants of London, yet a great deal is known about each of these classes. This is enough to show that definition by extension is not *necessary* to knowledge about a class. But when we come to consider infinite classes, we find that enumeration is not even theoretically possible for beings who only live for a finite time. We cannot enumerate all the natural numbers: they are 0, 1, 2, 3, *and so on*. At some point we must content ourselves with "and so on." We cannot enumerate all fractions or all irrational numbers, or all of any other infinite collection. Thus our knowledge in regard to all such collections can only be derived from a definition by intension.

These remarks are relevant, when we are seeking the definition of number, in three different ways. In the first place, numbers themselves form an infinite collection, and cannot therefore be defined by enumeration. In the second place, the collections having a given number of terms themselves presumably form an infinite collection: it is to be presumed, for example, that there are an infinite collection of trios in the world, for if this were not the case the total number of things in the world would be finite, which, though possible, seems unlikely. In the third place, we wish

to define "number" in such a way that infinite numbers may be possible; thus we must be able to speak of the number of terms in an infinite collection, and such a collection must be defined by intension, i.e. by a property common to all its members and peculiar to them.

For many purposes, a class and a defining characteristic of it are practically interchangeable. The vital difference between the two consists in the fact that there is only one class having a given set of members, whereas there are always many different characteristics by which a given class may be defined. Men may be defined as featherless bipeds, or as rational animals, or (more correctly) by the traits by which Swift delineates the Yahoos. It is this fact that a defining characteristic is never unique which makes classes useful; otherwise we could be content with the properties common and peculiar to their members.[3] Any one of these properties can be used in the place of the class whenever uniqueness is not important.

Returning now to the definition of number, it is clear that number is a way of bringing together certain collections, namely, those that have a given number of terms. We can suppose all couples in one bundle, all trios in another, and so on. In this way we obtain various bundles of collections, each bundle consisting of all the collections that have a certain number of terms. Each bundle is a class whose members are collections, i.e. classes; thus each is a class of classes. The bundle consisting of all couples, for example, is a class of classes: each couple is a class with two members, and the whole bundle of couples is a class with an infinite number of members, each of which is a class of two members.

How shall we decide whether two collections are to belong to the same bundle? The answer that suggests itself is: "Find out how many members each has, and put them in the same bundle if they have the same number of members." But this presupposes that we have defined numbers, and that we know how to discover how many terms a collection has. We are so used to the operation of counting that such a presupposition might easily pass unnoticed. In fact, however, counting, though familiar, is logically a very complex operation; moreover it is only available, as a means of discovering how many terms a collection has, when the collection is finite. Our definition of number must not assume in advance that all numbers are finite; and we cannot in any case, without a vicious circle, use counting to define numbers, because numbers are used in counting. We need, therefore, some other method of deciding when two collections have the same number of terms.

[3]As will be explained later, classes may be regarded as logical fictions, manufactured out of defining characteristics. But for the present it will simplify our exposition to treat classes as if they were real.

In actual fact, it is simpler logically to find out whether two collections have the same number of terms than it is to define what that number is. An illustration will make this clear. If there were no polygamy or polyandry anywhere in the world, it is clear that the number of husbands living at any moment would be exactly the same as the number of wives. We do not need a census to assure us of this, nor do we need to know what is the actual number of husbands and of wives. We know the number must be the same in both collections, because each husband has one wife and each wife has one husband. The relation of husband and wife is what is called "one-one."

A relation is said to be "one-one" when, if x has the relation in question to y, no other term x' has the same relation to y, and x does not have the same relation to any term y' other than y. When only the first of these two conditions is fulfilled, the relation is called "one-many"; when only the second is fulfilled, it is called "many-one." It should be observed that the number 1 is not used in these definitions.

In Christian countries, the relation of husband to wife is one-one; in Mahometan countries it is one-many; in Tibet it is many-one. The relation of father to son is one-many; that of son to father is many-one, but that of eldest son to father is one-one. If n is any number, the relation of n to $n+1$ is one-one; so is the relation of n to $2n$ or to $3n$. When we are considering only positive numbers, the relation of n to n^2 is one-one; but when negative numbers are admitted, it becomes two-one, since n and $-n$ have the same square. These instances should suffice to make clear the notions of one-one, one-many, and many-one relations, which play a great part in the principles of mathematics, not only in relation to the definition of numbers, but in many other connections.

Two classes are said to be "similar" when there is a one-one relation which correlates the terms of the one class each with one term of the other class, in the same manner in which the relation of marriage correlates husbands with wives. A few preliminary definitions will help us to state this definition more precisely. The class of those terms that have a given relation to something or other is called the *domain* of that relation: thus fathers are the domain of the relation of father to child, husbands are the domain of the relation of husband to wife, wives are the domain of the relation of wife to husband, and husbands and wives together are the domain of the relation of marriage. The relation of wife to husband is called the *converse* of the relation of husband to wife. Similarly *less* is the converse of *greater, later* is the converse of *earlier,* and so on. Generally, the converse of a given relation is that relation which holds between y and x whenever the given relation holds between x and y. The *converse domain* of a relation is the domain of its converse: thus the class of wives

is the converse domain of the relation of husband to wife. We may now state our definition of similarity as follows: –

> *One class is said to be "similar" to another when there is a one-one relation of which the one class is the domain, while the other is the converse domain.*

It is easy to prove (1) that every class is similar to itself, (2) that if a class α is similar to a class β, then β is similar to α, (3) that if α is similar to β and β to γ, then α is similar to γ. A relation is said to be *reflexive* when it possesses the first of these properties, *symmetrical* when it possesses the second, and *transitive* when it possesses the third. It is obvious that a relation which is symmetrical and transitive must be reflexive throughout its domain. Relations which possess these properties are an important kind, and it is worth while to note that similarity is one of this kind of relations.

It is obvious to common sense that two finite classes have the same number of terms if they are similar, but not otherwise. The act of counting consists in establishing a one-one correlation between the set of objects counted and the natural numbers (excluding 0) that are used up in the process. Accordingly common sense concludes that there are as many objects in the set to be counted as there are numbers up to the last number used in the counting. And we also know that, so long as we confine ourselves to finite numbers, there are just n numbers from 1 up to n. Hence it follows that the last number used in counting a collection is the number of terms in the collection, provided the collection is finite. But this result, besides being only applicable to finite collections, depends upon and assumes the fact that two classes which are similar have the same number of terms; for what we do when we count (say) 10 objects is to show that the set of these objects is similar to the set of numbers 1 to 10. The notion of similarity is logically presupposed in the operation of counting, and is logically simpler though less familiar. In counting, it is necessary to take the objects counted in a certain order, as first, second, third, etc., but order is not of the essence of number: it is an irrelevant addition, an unnecessary complication from the logical point of view. The notion of similarity does not demand an order: for example, we saw that the number of husbands is the same as the number of wives, without having to establish an order of precedence among them. The notion of similarity also does not require that the classes which are similar should be finite. Take, for example, the natural numbers (excluding 0) on the one hand, and the fractions which have 1 for their numerator on the other hand: it is obvious that we can correlate 2 with 1/2, 3 with 1/3, and so on, thus proving that the two classes are similar.

We may thus use the notion of "similarity" to decide when two collections are to belong to the same bundle, in the sense in which we were asking this question earlier in this chapter. We want to make one bundle containing the class that has no members: this will be for the number 0. Then we want a bundle of all the classes that have one member: this will be for the number 1. Then, for the number 2, we want a bundle consisting of all couples; then one of all trios; and so on. Given any collection, we can define the bundle it is to belong to as being the class of all those collections that are "similar" to it. It is very easy to see that if (for example) a collection has three members, the class of all those collections that are similar to it will be the class of trios. And whatever number of terms a collection may have, those collections that are "similar" to it will have the same number of terms. We may take this as a *definition* of "having the same number of terms." It is obvious that it gives results conformable to usage so long as we confine ourselves to finite collections.

So far we have not suggested anything in the slightest degree paradoxical. But when we come to the actual definition of numbers we cannot avoid what must at first sight seem a paradox, though this impression will soon wear off. We naturally think that the class of couples (for example) is something different from the number 2. But there is no doubt about the class of couples: it is indubitable and not difficult to define, whereas the number 2, in any other sense, is a metaphysical entity about which we can never feel sure that it exists or that we have tracked it down. It is therefore more prudent to content ourselves with the class of couples, which we are sure of, than to hunt for a problematical number 2 which must always remain elusive. Accordingly we set up the following definition: -

The number of a class is the class of all those classes that are similar to it.

Thus the number of a couple will be the class of all couples. In fact, the class of all couples will *be* the number 2, according to our definition. At the expense of a little oddity, this definition secures definiteness and indubitableness; and it is not difficult to prove that numbers so defined have all the properties that we expect numbers to have.

We may now go on to define numbers in general as any one of the bundles into which similarity collects classes. A number will be a set of classes such as that any two are similar to each other, and none outside the set are similar to any inside the set. In other words, a number (in general) is any collection which is the number of one of its members; or, more simply still:

A number is anything which is the number of some class.

Such a definition has a verbal appearance of being circular, but in fact it is not. We define "the number of a given class" without using the notion of number in general; therefore we may define number in general in terms of "the number of a given class" without committing any logical error.

Definitions of this sort are in fact very common. The class of fathers, for example, would have to be defined by first defining what it is to be the father of somebody; then the class of fathers will be all those who are somebody's father. Similarly if we want to define square numbers (say), we must first define what we mean by saying that one number is the square of another, and then define square numbers as those that are the squares of other numbers. This kind of procedure is very common, and it is important to realize that it is legitimate and even often necessary.

We have now given a definition of numbers which will serve for finite collections. It remains to be seen how it will serve for infinite collections. But first we must decide what we mean by "finite" and "infinite," which cannot be done within the limits [here].

III. Mathematics and logic

Mathematics and logic, historically speaking, have been entirely distinct studies. Mathematics has been connected with science, logic with Greek. But both have developed in modern times: logic has become more mathematical and mathematics has become more logical. The consequence is that it has now become wholly impossible to draw a line between the two; in fact, the two are one. They differ as boy and man: logic is the youth of mathematics and mathematics is the manhood of logic. This view is resented by logicians who, having spent their time in the study of classical texts, are incapable of following a piece of symbolic reasoning, and by mathematicians who have learnt a technique without troubling to inquire into its meaning or justification. Both types are now fortunately growing rarer. So much of modern mathematical work is obviously on the borderline of logic, so much of modern logic is symbolic and formal, that the very close relationship of logic and mathematics has become obvious to every instructed student. The proof of their identity is, of course, a matter of detail: starting with premisses which would be universally admitted to belong to logic, and arriving by deduction at results which as obviously belong to mathematics, we find that there is no point at which a sharp line can be drawn, with logic to the left and mathematics to the right. If there are still those who do not admit the identity of logic and mathematics, we may challenge them to indicate at what point, in the successive definitions and deductions of *Principia Mathematica*, they consider that logic ends and mathematics begins. It will then be obvious that any answer must be quite arbitrary.

In the earlier chapters of this book, starting from the natural numbers, we have first defined "cardinal number" and shown how to generalise the conception of number, and have then analysed the conceptions involved in the definition, until we found ourselves dealing with the fundamentals of logic. In a synthetic, deductive treatment these fundamentals come first, and the natural numbers are only reached after a long journey. Such treatment, though formally more correct than that which we have adopted, is more difficult for the reader, because the ultimate logical concepts and propositions with which it starts are remote and unfamiliar as compared with the natural numbers. Also they represent the present frontier of knowledge, beyond which is the still unknown; and the dominion of knowledge over them is not as yet very secure.

It used to be said that mathematics is the science of "quantity." "Quantity" is a vague word, but for the sake of argument we may replace it by the word "number." The statement that mathematics is the science of number would be untrue in two different ways. On the one hand, there are recognised branches of mathematics which have nothing to do with number – all geometry that does not use co-ordinates or measurement, for example: projective and descriptive geometry, down to the point at which co-ordinates are introduced, does not have to do with number, or even with quantity in the sense of *greater* or *less*. On the other hand, through the definition of cardinals, through the theory of induction and ancestral relations, through the general theory of series, and through the definitions of the arithmetical operations, it has become possible to generalize much that used to be proved only in connection with numbers. The most elementary properties of numbers are concerned with one-one relations, and similarity between classes. Addition is concerned with the construction of mutually exclusive classes respectively similar to a set of classes which are not known to be mutually exclusive. Multiplication is merged in the theory of "selections," i.e. of a certain kind of one-many relations. Finitude is merged in the general study of ancestral relations, which yields the whole theory of mathematical induction. The ordinal properties of the various kinds of number-series, and the elements of the theory of continuity of functions and the limits of functions, can be generalised so as no longer to involve any essential reference to numbers. It is a principle, in all formal reasoning, to generalize to the utmost, since we thereby secure that a given process of deduction shall have more widely applicable results; we are, therefore, in thus generalizing the reasoning of arithmetic, merely following a precept which is universally admitted in mathematics. And in thus generalizing we have, in effect,

created a set of new deductive systems, in which traditional arithmetic is at once dissolved and enlarged; but whether any one of these new deductive systems – for example, the theory of selections – is to be said to belong to logic or to arithmetic is entirely arbitrary, and incapable of being decided rationally.

We are thus brought face to face with the question: What is the subject, which may be called indifferently either mathematics or logic? Is there any way in which we can define it?

Certain characteristics of the subject are clear. To begin with, we do not, in this subject, deal with particular things or particular properties: we deal formally with what can be said about *any* thing or *any* property. We are prepared to say that one and one are two, but not that Socrates and Plato are two, because, in our capacity of logicians or pure mathematicians, we have never heard of Socrates and Plato. A world in which there were no such individuals would still be a world in which one and one are two. It is not open to us, as pure mathematicians or logicians, to mention anything at all, because, if we do so, we introduce something irrelevant and not formal. We may make this clear by applying it to the case of the syllogism. Traditional logic says: "All men are mortal, Socrates is a man, therefore Socrates is mortal." Now it is clear that what we *mean* to assert, to begin with, is only that the premisses imply the conclusion, not that premisses and conclusion are actually true; even the most traditional logic points out that the actual truth of the premisses is irrelevant to logic. Thus the first change to be made in the above traditional syllogism is to state it in the form: "If all men are mortal and Socrates is a man, then Socrates is mortal." We may now observe that it is intended to convey that this argument is valid in virtue of its *form*, not in virtue of the particular terms occurring in it. If we had omitted "Socrates is a man" from our premisses, we should have had a non-formal argument, only admissible because Socrates is in fact a man; in that case we could not have generalized the argument. But when, as above, the argument is *formal*, nothing depends upon the terms that occur in it. Thus we may substitute α for *men*, β for *mortals*, and x for Socrates, where α and β are any classes whatever, and x is any individual. We then arrive at the statement: "No matter what possible values x and α and β may have, if all α's are β's and x is an α, then x is a β"; in other words, "the propositional function 'if all α's are β and x is an α, then x is a β' is always true." Here at last we have a proposition of logic – the one which is only *suggested* by the traditional statement about Socrates and men and mortals.

It is clear that, if *formal* reasoning is what we are aiming at, we shall always arrive ultimately at statements like the above, in which no actual things or properties are mentioned; this will happen through the mere desire not to waste our time proving in a particular case what can be proved generally. It would be ridiculous to go through a long argument about Socrates, and then go through precisely the same argument again about Plato. If our argument is one (say) which holds of all men, we shall prove it concerning "x," with the hypothesis "if x is a man." With this hypothesis, the argument will retain its hypothetical validity even when x is not a man. But now we shall find that our argument would still be valid if, instead of supposing x to be a man, we were to suppose him to be a monkey or a goose or a Prime Minister. We shall therefore not waste our time taking as our premiss "x is a man" but shall take "x is an α," where α is any class of individuals, or "ϕx" where ϕ is any propositional function of some assigned type. Thus the absence of all mention of particular things or properties in logic or pure mathematics is a necessary result of the fact that this study is, as we say, "purely formal."

At this point we find ourselves faced with a problem which is easier to state than to solve. The problem is: "What are the constituents of a logical proposition?" I do not know the answer, but I propose to explain how the problem arises.

Take (say) the proposition "Socrates was before Aristotle." Here it seems obvious that we have a relation between two terms, and that the constituents of the proposition (as well as of the corresponding fact) are simply the two terms and the relation, i.e. Socrates, Aristotle, and *before*. (I ignore the fact that Socrates and Aristotle are not simple; also the fact that what appear to be their names are really truncated descriptions. Neither of these facts is relevant to the present issue.) We may represent the general form of such propositions by "xRy," which may be read "x has the relation R to y." This general form may occur in logical propositions, but no particular instance of it can occur. Are we to infer that the general form itself is a constituent of such logical propositions?

Given a proposition, such as "Socrates is before Aristotle," we have certain constituents and also a certain form. But the form is not itself a new constituent; if it were, we should need a new form to embrace both it and the other constituents. We can, in fact, turn *all* the constituents of a proposition into variables, while keeping the form unchanged. This is what we do when we use such a schema as "xRy," which stands for any one of a certain class of propositions, namely, those asserting relations between two terms. We can proceed to general assertions, such as "xRy is sometimes true" - i.e. there are cases where dual relations hold. This assertion will belong to logic (or mathematics) in the sense in which we are using the word. But in this assertion we do not mention any particular things or particular relations; no particular things or relations can

ever enter into a proposition of pure logic. We are left with pure *forms* as the only possible constituents of logical propositions.

I do not wish to assert positively that pure forms - e.g. the form "xRy" - do actually enter into propositions of the kind we are considering. The question of the analysis of such propositions is a difficult one, with conflicting considerations on the one side and on the other. We cannot embark upon this question now, but we may accept, as a first approximation, the view that *forms* are what enter into logical propositions as their constituents. And we may explain (though not formally define) what we mean by the "form" of a proposition as follows: -

> The "form" of a proposition is that, in it, that remains unchanged when every constituent of the proposition is replaced by another.

Thus "Socrates is earlier than Aristotle" has the same form as "Napoleon is greater than Wellington," though every constituent of the two propositions is different.

We may thus lay down, as a necessary (though not sufficient) characteristic of logical or mathematical propositions, that they are to be such as can be obtained from a proposition containing no variables (i.e. no such words as *all, some, a, the,* etc.) by turning every constituent into a variable and asserting that the result is always true or sometimes true, or that it is always true in respect of some of the variables that the result is sometimes true in respect of the others, or any variant of these forms. And another way of stating the same thing is to say that logic (or mathematics) is concerned only with *forms*, and is concerned with them only in the way of stating that they are always or sometimes true - with all the permutations or "always" and "sometimes" that may occur.

There are in every language some words whose sole function is to indicate form. These words, broadly speaking, are commonest in languages having fewest inflections. Take "Socrates is human." Here "is" is not a constituent of the proposition, but merely indicates the subject-predicate form. Similarly in "Socrates is earlier than Aristotle," "is" and "than" merely indicate form; the proposition is the same as "Socrates precedes Aristotle," in which these words have disappeared and the form is otherwise indicated. Form, as a rule, *can* be indicated otherwise than by specific words: the order of the words can do most of what is wanted. But this principle must not be pressed. For example, it is difficult to see how we could conveniently express molecular forms of propositions (i.e. what we call "truth-functions") without any word at all. We saw ... that one word is enough for this purpose, namely, a word or symbol expressing *incompatibility*. But without even one we should find ourselves in difficulties. This, however, is not the point that is important for our present purpose. What is important for us is to observe that form may be the one concern of a general proposition, even when no word or symbol in that proposition designates the form. If we wish to speak about the form itself, we must have a word for it; but if, as in mathematics, we wish to speak about all propositions that have the form, a word for the form will usually be found not indispensable; probably in theory it is *never* indispensable.

Assuming – as I think we may – that the forms of propositions *can* be represented by the forms of the propositions in which they are expressed without any special word for forms, we should arrive at a language in which everything formal belonged to syntax and not to vocabulary. In such a language we could express *all* the propositions of mathematics even if we did not know one single word of the language. The language of mathematical logic, if it were perfected, would be such a language. We should have symbols for variables, such as "x" and "R" and "y," arranged in various ways; and the way of arrangement would indicate that something was being said to be true of all values or some values of the variables. We should not need to know any words, because they would only be needed for giving values to the variables, which is the business of the applied mathematician, not of the pure mathematician or logician. It is one of the marks of a proposition of logic that, given a suitable language, such a proposition can be asserted in such a language by a person who knows the syntax without knowing a single word of the vocabulary.

But, after all, there are words that express form, such as "is" and "than." And in every symbolism hitherto invented for mathematical logic there are symbols having constant formal meanings. We may take as an example the symbol for incompatibility which is employed in building up truth-functions. Such words or symbols may occur in logic. The question is: How are we to define them?

Such words or symbols express what are called "logical constants." Logical constants may be defined exactly as we defined forms; in fact, they are in essence the same thing. A fundamental logical constant will be that which is in common among a number of propositions, any one of which can result from any other by substitution of terms one for another. For example, "Napoleon is greater than Wellington" results from "Socrates is earlier than Aristotle" by the substitution of "Napoleon" for "Socrates," "Wellington" for "Aristotle," and "greater" for "earlier." Some propositions can be obtained in this way from the prototype "Socrates is earlier than Aristotle" and some cannot; those that can are those that are of the form "xRy," i.e. express dual relations. We cannot obtain from the above prototype by term-for-term substitution such

propositions as "Socrates is human" or "the Athenians gave the hemlock to Socrates," because the first is of the subject-predicate form and the second expresses a three-term relation. If we are to have any words in our pure logical language, they must be such as express "logical constants," and "logical constants" will always either be, or be derived from, what is in common among a group of propositions derivable from each other, in the above manner, by term-for-term substitution. And this which is in common is what we call "form."

In this sense all the "constants" that occur in pure mathematics are logical constants. The number 1, for example, is derivative from propositions of the form: "There is a term c such that ϕx is true when, and only when, x is c." This is a function of ϕ, and various different propositions result from giving different values to ϕ. We may (with a little omission of intermediate steps not relevant to our present purpose) take the above function of ϕ as what is meant by "the class determined by ϕ is a unit class" or "the class determined by ϕ is a member of 1" (1 being a class of classes). In this way, propositions in which 1 occurs acquire a meaning which is derived from a certain constant logical form. And the same will be found to be the case with all mathematical constants: all are logical constants, or symbolic abbeviations whose full use in a proper context is defined by means of logical constants.

But although all logical (or mathematical) propositions can be expressed wholly in terms of logical constants together with variables, it is not the case that, conversely, all propositions that can be expressed in this way are logical. We have found so far a necessary but not a sufficient criterion of mathematical propositions. We have sufficiently defined the character of the primitive *ideas* in terms of which all the ideas of mathematics can be *defined*, but not of the primitive *propositions* from which all the propositions of mathematics can be *deduced*. This is a more difficult matter, as to which it is not yet known what the full answer is.

We may take the axiom of infinity as an example of a proposition which, though it can be enunciated in logical terms, cannot be asserted by logic to be true. All the propositions of logic have a characteristic which used to be expressed by saying that they were analytic, or that their contradictories were self-contradictory. This mode of statement, however, is not satisfactory. The law of contradiction is merely one among logical propositions; it has no special pre-eminence; and the proof that the contradictory of some proposition is self-contradictory is likely to require other principles of deduction besides the law of contradiction. Nevertheless, the characteristic of logical propositions that we are in search of is the one which was felt, and intended to be defined, by those who said that it consisted in deducibility from the law of contradiction.

This characteristic, which, for the moment, we may call *tautology,* obviously does not belong to the assertion that the number of individuals in the universe is n, whatever number n may be. But for the diversity of types, it would be possible to prove logically that there are classes of n terms, where n is any finite integer; or even that there are classes of $\aleph_0$ terms. But, owing to types, such proofs . . . are fallacious. We are left to empirical observation to determine whether there are as many as n individuals in the world. Among "possible" worlds, in the Leibnizian sense, there will be worlds having one, two, three, . . . individuals. There does not even seem any logical necessity why there should be even one individual[4] – why, in fact, there should be any world at all. The ontological proof of the existence of God, if it were valid, would establish the logical necessity of at least one individual. But it is generally recognized as invalid, and in fact rests upon a mistaken view of existence – i.e. it fails to realize that existence can only be asserted of something described, not of something named, so that it is meaningless to argue from "this is the so-and-so" and "the so-and-so exists" to "this exists." If we reject the ontological argument, we seem driven to conclude that the existence of a world is an accident – i.e. it is not logically necessary. If that be so, no principle of logic can assert "existence" except under a hypothesis, i.e. none can be of the form "the propositional function so-and-so is sometimes true." Propositions of this form, when they occur in logic, will have to occur as hypotheses or consequences of hypotheses, not as complete asserted propositions. The complete asserted propositions of logic will all be such as affirm that some propositional function is *always* true. For example, it is always true that if p implies q and q implies r then p implies r, or that, if all α's are β's and x is an α then x is a β. Such propositions may occur in logic, and their truth is independent of the existence of the universe. We may lay it down that, if there were no universe, *all* general propositions would be true; for the contradictory of a general proposition . . . is a proposition asserting existence, and would therefore always be false if no universe existed.

Logical propositions are such as can be known *a priori,* without study of the actual world. We only know from a study of empirical facts that Socrates is a man, but we know the correctness of the syllogism in its abstract form (i.e. when it is stated in terms of variables) without needing any appeal to experience. This is a characteristic, not of logical propositions in themselves, but of the way in which we know them. It has, however, a bearing upon the question what their nature may be, since there

[4]The primitive propositions in *Principia Mathematica* are such as to allow the inference that at least one individual exists. But I now view this as a defect in logical purity.

are some kinds of propositions which it would be very difficult to suppose we could know without experience.

It is clear that the definition of "logic" or "mathematics" must be sought by trying to give a new definition of the old notion of "analytic" propositions. Although we can no longer be satisfied to define logical propositions as those that follow from the law of contradiction, we can and must still admit that they are a wholly different class of propositions from those that we come to know empirically. They all have the characteristic which, a moment ago, we agreed to call "tautology." This, combined with the fact that they can be expressed wholly in terms of variables and logical constants (a logical constant being something which remains constant in a proposition even when *all* its constituents are changed), will give the definition of logic or pure mathematics. For the moment, I do not know how to define "tautology."[5] It would be easy to offer a definition which might seem satisfactory for a while; but I know of none that I feel to be satisfactory, in spite of feeling thoroughly familiar with the characteristic of which a definition is wanted. At this point, therefore, for the moment, we reach the frontier of knowledge on our backward journey into the logical foundations of mathematics.

We have now come to an end of our somewhat summary introduction to mathematical philosophy. It is impossible to convey adequately the ideas that are concerned in this subject so long as we abstain from the use of logical symbols. Since ordinary language has no words that naturally express exactly what we wish to express, it is necessary, so long as we adhere to ordinary language, to strain words into unusual meanings; and the reader is sure, after a time if not at first, to lapse into attaching the usual meanings to words, thus arriving at wrong notions as to what is intended to be said. Moreover, ordinary grammar and syntax is extraordinarily misleading. This is the case, e.g. as regards numbers; "ten men" is grammatically the same form as "white men," so that 10 might be thought to be an adjective qualifying "men." It is the case, again, wherever propositional functions are involved, and in particular as regards existence and descriptions. Because language is misleading, as well as because it is diffuse and inexact when applied to logic (for which it was never intended), logical symbolism is absolutely necessary to any exact or thorough treatment of our subject. Those readers, therefore, who wish to acquire a mastery of the principles of mathematics, will, it is to be hoped, not shrink from the labour of mastering the symbols –

[5]The importance of "tautology" for a definition of mathematics was pointed out to me by my former pupil Ludwig Wittgenstein, who was working on the problem. I do not know whether he has solved it, or even whether he is alive or dead.

a labour which is, in fact, much less than might be thought. As the above hasty survey must have made evident, there are innumerable unsolved problems in the subject, and much work needs to be done. If any student is led into a serious study of mathematical logic by this little book, it will have served the chief purpose for which it has been written.

On the infinite

DAVID HILBERT

As a result of his penetrating critique, Weierstrass has provided a solid foundation for mathematical analysis. By elucidating many notions, in particular those of minimum, function, and differential quotient, he removed the defects which were still found in the infinitesimal calculus, rid it of all confused notions about the infinitesimal, and thereby completely resolved the difficulties which stem from that concept. If in analysis today there is complete agreement and certitude in employing the deductive methods which are based on the concepts of irrational number and limit, and if in even the most complex questions of the theory of differential and integral equations, notwithstanding the use of the most ingenious and varied combinations of the different kinds of limits, there nevertheless is unanimity with respect to the results obtained, then this happy state of affairs is due primarily to Weierstrass's scientific work.

And yet in spite of the foundation Weierstrass has provided for the infinitesimal calculus, disputes about the foundations of analysis still go on.

These disputes have not terminated because the meaning of the *infinite,* as that concept is used in mathematics, has never been completely clarified. Weierstrass's analysis did indeed eliminate the infinitely large and the infinitely small by reducing statements about them to [statements about] relations between finite magnitudes. Nevertheless the infinite still appears in the infinite numerical series which defines the real numbers and in the concept of the real number system which is thought of as a completed totality existing all at once.

In his foundation for analysis, Weierstrass accepted unreservedly and used repeatedly those forms of logical deduction in which the concept of the infinite comes into play, as when one treats of *all* real numbers with a certain property or when one argues that *there exist* real numbers with a certain property.

Delivered June 4, 1925, before a congress of the Westphalian Mathematical Society in Munster, in honor of Karl Weierstrass. Translated by Erna Putnam and Gerald J. Massey from *Mathematische Annalen* (Berlin) vol. 95 (1926), pp. 161–90. Permission for the translation and inclusion of the article in this volume was kindly granted by the publishers, Springer Verlag.

Hence the infinite can reappear in another guise in Weierstrass's theory and thus escape the precision imposed by his critique. It is, therefore, *the problem of the infinite* in the sense just indicated which we need to resolve once and for all. Just as in the limit processes of the infinitesimal calculus, the infinite in the sense of the infinitely large and the infinitely small proved to be merely a figure of speech, so too we must realize that the infinite in the sense of an infinite totality, where we still find it used in deductive methods, is an illusion. Just as operations with the infinitely small were replaced by operations with the finite which yielded exactly the same results and led to exactly the same elegant formal relationships, so in general must deductive methods based on the infinite be replaced by finite procedures which yield exactly the same results; i.e., which make possible the same chains of proofs and the same methods of getting formulas and theorems.

The goal of my theory is to establish once and for all the certitude of mathematical methods. This is a task which was not accomplished even during the critical period of the infinitesimal calculus. This theory should thus complete what Weierstrass hoped to achieve by his foundation for analysis and toward the accomplishment of which he has taken a necessary and important step.

But a still more general perspective is relevant for clarifying the concept of the infinite. A careful reader will find that the literature of mathematics is glutted with inanities and absurdities which have had their source in the infinite. For example, we find writers insisting, as though it were a restrictive condition, that in rigorous mathematics only a *finite* number of deductions are admissible in a proof – as if someone had succeeded in making an infinite number of them.

Also old objections which we supposed long abandoned still reappear in different forms. For example, the following recently appeared: Although it may be possible to introduce a concept without risk, i.e., without getting contradictions, and even though one can prove that its introduction causes no contradictions to arise, still the introduction of the concept is not thereby justified. Is not this exactly the same objection which was once brought against complex-imaginary numbers when it was said: "True, their use doesn't lead to contradictions. Nevertheless their introduction is unwarranted, for imaginary magnitudes do not exist"? If, apart from proving consistency, the question of the justification of a measure is to have any meaning, it can consist only in ascertaining whether the measure is accompanied by commensurate success. Such success is in fact essential, for in mathematics as elsewhere success is the supreme court to whose decisions everyone submits.

As some people see ghosts, another writer seems to see contradictions even where no statements whatsoever have been made, viz., in the concrete world of sensation, the "consistent functioning" of which he takes as special assumption. I myself have always supposed that only statements, and hypotheses insofar as they lead through deductions to statements, could contradict one another. The view that facts and events could themselves be in contradiction seems to me to be a prime example of careless thinking.

The foregoing remarks are intended only to establish the fact that the definitive clarification of *the nature of the infinite,* instead of pertaining just to the sphere of specialized scientific interests, is needed for *the dignity of the human intellect* itself.

From time immemorial, the infinite has stirred men's *emotions* more than any other question. Hardly any other *idea* has stimulated the mind so fruitfully. Yet, no other *concept* needs *clarification* more than it does.

Before turning to the task of clarifying the nature of the infinite, we should first note briefly what meaning is actually given to the infinite. First let us see what we can learn from physics. One's first naïve impression of natural events and of matter is one of permanency, of continuity. When we consider a piece of metal or a volume of liquid, we get the impression that they are unlimitedly divisible, that their smallest parts exhibit the same properties that the whole does. But wherever the methods of investigating the physics of matter have been sufficiently refined, scientists have met divisibility boundaries which do not result from the shortcomings of their efforts but from the very nature of things. Consequently we could even interpret the tendency of modern science as emancipation from the infinitely small. Instead of the old principle *natura non facit saltus,* we might even assert the opposite, viz., "nature makes jumps."

It is common knowledge that all matter is composed of tiny building blocks called "atoms," the combinations and connections of which produce all the variety of macroscopic objects. Still physics did not stop at the atomism of matter. At the end of the last century there appeared the atomism of electricity which seems much more bizarre at first sight. Electricity, which until then had been thought of as a fluid and was considered the model of a continuously active agent, was then shown to be built up of positive and negative *electrons.*

In addition to matter and electricity, there is one other entity in physics for which the law of conservation holds, viz., energy. But it has been established that even energy does not unconditionally admit of infinite divisibility. Planck has discovered *quanta of energy.*

Hence, a homogeneous continuum which admits of the sort of divisibility needed to realize the infinitely small is nowhere to be found in reality. The infinite divisibility of a continuum is an operation which exists only in thought. It is merely an idea which is in fact impugned by the results of our observations of nature and of our physical and chemical experiments.

The second place where we encounter the question of whether the infinite is found in nature is in the consideration of the universe as a whole. Here we must consider the expanse of the universe to determine whether it embraces anything infinitely large. But here again modern science, in particular astronomy, has reopened the question and is endeavoring to solve it, not by the defective means of metaphysical speculation, but by reasons which are based on experiment and on the application of the laws of nature. Here, too, serious objections against infinity have been found. *Euclidean* geometry necessarily leads to the postulate that space is infinite. Although euclidean geometry is indeed a consistent conceptual system, it does not thereby follow that euclidean geometry actually holds in reality. Whether or not real space is euclidean can be determined only through observation and experiment. The attempt to prove the infinity of space by pure speculation contains gross errors. From the fact that outside a certain portion of space there is always more space, it follows only that space is unbounded, not that it is infinite. Unboundedness and finiteness are compatible. In so-called *elliptical* geometry, mathematical investigation furnishes the natural model of a finite universe. Today the abandonment of euclidean geometry is no longer merely a mathematical or philosophical speculation but is suggested by considerations which originally had nothing to do with the question of the finiteness of the universe. Einstein has shown that euclidean geometry must be abandoned. On the basis of his gravitational theory, he deals with cosmological questions and shows that a finite universe is possible. Moreover, all the results of astronomy are perfectly compatible with the postulate that the universe is elliptical.

We have established that the universe is finite in two respects, i.e., as regards the infinitely small and the infinitely large. But it may still be the case that the infinite occupies a justified place *in our thinking*, that it plays the role of an indispensable concept. Let us see what the situation is in mathematics. Let us first interrogate that purest and simplest offspring of the human mind, viz., number theory. Consider one formula out of the rich variety of elementary formulas of number theory, e.g., the formula

$$1^2+2^2+3^2\cdots+n^2=\tfrac{1}{6}n(n+1)(2n+1)$$

Since we may substitute any integer whatsoever for n, for example $n=2$ or $n=5$, this formula implicitly contains *infinitely many* propositions. This characteristic is essential to a formula. It enables the formula to represent the solution of an arithmetical problem and necessitates a special idea for its proof. On the other hand, the individual numerical equations

$$1^2+2^2=\tfrac{1}{6}\cdot 2\cdot 3\cdot 5$$
$$1^2+2^2+3^2+4^2+5^2=\tfrac{1}{6}\cdot 5\cdot 6\cdot 11$$

can be verified simply by calculation and hence individually are of no especial interest.

We encounter a completely different and quite unique conception of the notion of infinity in the important and fruitful method of *ideal elements*. The method of ideal elements is used even in elementary plane geometry. The points and straight lines of the plane originally are real, actually existent objects. One of the axioms that hold for them is the axiom of connection: one and only one straight line passes through two points. It follows from this axiom that two straight lines intersect at most at one point. There is no theorem that two straight lines always intersect at some point, however, for the two straight lines might well be parallel. Still we know that by introducing ideal elements, viz., infinitely long lines and points at infinity, we can make the theorem that two straight lines always intersect at one and only one point come out universally true. These ideal "infinite" elements have the advantage of making the system of connection laws as simple and perspicuous as possible. Moreover, because of the symmetry between a point and a straight line, there results the very fruitful principle of duality for geometry.

Another example of the use of ideal elements are the familiar *complex-imaginary* magnitudes of algebra which serve to simplify theorems about the existence and number of the roots of an equation.

Just as infinitely many straight lines, viz., those parallel to each other, are used to define an ideal point in geometry, so certain systems of infinitely many numbers are used to define an *ideal number*. This application of the principle of ideal elements is the most ingenious of all. If we apply this principle systematically throughout an algebra, we obtain exactly the same simple and familiar laws of division which hold for the familiar whole numbers 1, 2, 3, 4,.... We are already in the domain of higher arithmetic.

We now come to the most aesthetic and delicately erected structure of mathematics, viz., analysis. You already know that infinity plays the leading role in analysis. In a certain sense, mathematical analysis is a symphony of the infinite.

The tremendous progress made in the infinitesimal calculus results mainly from operating with mathematical systems of infinitely many elements. But, as it seemed very plausible to identify the infinite with the "very large", there soon arose inconsistencies which were known in part to the ancient sophists, viz., the so-called paradoxes of the infinitesimal calculus. But the recognition that many theorems which hold for the finite (for example, the part is smaller than the whole, the existence of a minimum and a maximum, the interchangeability of the order of the terms of a sum or a product) cannot be immediately and unrestrictedly extended to the infinite, marked fundamental progress. I said at the beginning of this paper that these questions have been completely clarified, notably through Weierstrass's acuity. Today, analysis is not only infallible within its domain but has become a practical instrument for using the infinite.

But analysis alone does not provide us with the deepest insight into the nature of the infinite. This insight is procured for us by a discipline which comes closer to a general philosophical way of thinking and which was designed to cast new light on the whole complex of questions about the infinite. This discipline, created by George Cantor, is set theory. In this paper, we are interested only in that unique and original part of set theory which forms the central core of Cantor's doctrine, viz., the theory of *transfinite* numbers. This theory is, I think, the finest product of mathematical genius and one of the supreme achievements of purely intellectual human activity. What, then, is this theory?

Someone who wished to characterize briefly the new conception of the infinite which Cantor introduced might say that in analysis we deal with the infinitely large and the infinitely small only as limiting concepts, as something becoming, happening, i.e., with the *potential infinite*. But this is not the true infinite. We meet the true infinite when we regard the totality of numbers $1, 2, 3, 4, \ldots$ itself as a completed unity, or when we regard the points of an interval as a totality of things which exists all at once. This kind of infinity is known as *actual infinity*.

Frege and Dedekind, the two mathematicians most celebrated for their work in the foundations of mathematics, independently of each other used the actual infinite to provide a foundation for arithmetic which was independent of both intuition and experience. This foundation was based solely on pure logic and made use only of deductions that were purely logical. Dedekind even went so far as not to take the notion of finite number from intuition but to derive it logically by employing the concept of an infinite set. But is was Cantor who systematically developed the concept of the actual infinite. Consider the two examples of the infinite already mentioned

1. $1, 2, 3, 4, \ldots$.
2. The points of the interval 0 to 1 or, what comes to the same thing, the totality of real numbers between 0 and 1.

It is quite natural to treat these examples from the point of view of their size. But such a treatment reveals amazing results with which every mathematician today is familiar. For when we consider the set of all rational numbers, i.e., the fractions $1/2, 1/3, 2/3, 1/4, \ldots, 3/7, \ldots$, we notice that – from the sole standpoint of its size – this set is no larger than the set of integers. Hence we say that the rational numbers can be counted in the usual way; i.e., that they are enumerable. The same holds for the set of all roots of numbers, indeed even for the set of all algebraic numbers. The second example is analogous to the first. Surprisingly enough, the set of all the points of a square or cube is no larger than the set of points of the interval 0 to 1. Similarly for the set of all continuous functions. On learning these facts for the first time, you might think that from the point of view of size there is only one unique infinite. No, indeed! The sets in examples (1) and (2) are not, as we say, "equivalent". Rather, the set (2) cannot be enumerated, for it is larger than the set (1). We meet what is new and characteristic in Cantor's theory at this point. The points of an interval cannot be counted in the usual way, i.e., by counting $1, 2, 3, \ldots$. But, since we admit the actual infinite, we are not obliged to stop here. When we have counted $1, 2, 3, \ldots$, we can regard the objects thus enumerated as an infinite set existing all at once in a particular order. If, following Cantor, we call the type of this order ω, then counting continues naturally with $\omega+1, \omega+2, \ldots$ up to $\omega+\omega$ or $\omega\cdot 2$, and then again

$$(\omega\cdot 2)+1,\ (\omega\cdot 2)+2,\ (\omega\cdot 2)+3, \ldots\ (\omega\cdot 2)+\omega \text{ or } \omega\cdot 3,$$

and further

$$\omega\cdot 2, \omega\cdot 3, \omega\cdot 4, \ldots, \omega\cdot\omega \text{ (or } \omega^2), \omega^2+1, \ldots,$$

so that we finally get this table:

$$\begin{array}{l}
1, 2, 3, \ldots \\
\omega, \omega+1, \omega+2, \ldots \\
\omega\cdot 2, (\omega\cdot 2)+1, (\omega\cdot 2)+2, \ldots \\
\omega\cdot 3, (\omega\cdot 3)+1, (\omega\cdot 3)+2, \ldots \\
\vdots \\
\omega^2, \omega^2+1, \ldots \\
\omega^2+\omega, \omega^2+\omega\cdot 2, \omega^2+\omega\cdot 3, \ldots \\
\omega^2\cdot 2, (\omega^2\cdot 2)+1, \ldots
\end{array}$$

$$(\omega^2 \cdot 2) + \omega,\ (\omega^2 \cdot 2) + (\omega \cdot 2), \ldots$$
$$\omega^3, \ldots$$
$$\omega^4, \ldots$$
$$\vdots$$
$$\omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \ldots$$

These are Cantor's first transfinite numbers, or, as he called them, the numbers of the second number class. We arrive at them simply by extending counting beyond the ordinarily enumerably infinite, i.e., by a natural and uniquely determined consistent continuation of ordinary finite counting. As until now we counted only the first, second, third,... member of a set, we not count also the ωth, $(\omega+1)$st, ..., ω^ωth member.

Given these developments one naturally wonders whether or not, by using these transfinite numbers, one can really count those sets which cannot be counted in the ordinary way.

On the basis of these concepts, Cantor developed the theory of transfinite numbers quite successfully and invented a full calculus for them. Thus, thanks to the Herculean collaboration of Frege, Dedekind, and Cantor, the infinite was made king and enjoyed a reign of great triumph. In daring flight, the infinite had reached a dizzy pinnacle of success.

But reaction was not lacking. It took in fact a very dramatic form. It set in perfectly analogously to the way reaction had set in against the development of the infinitesimal calculus. In the joy of discovering new and important results, mathematicians paid too little attention to the validity of their deductive methods. For, simply as a result of employing definitions and deductive methods which had become customary, contradictions began gradually to appear. These contradictions, the so-called paradoxes of set theory, though at first scattered, became progressively more acute and more serious. In particular, a contradiction discovered by Zermelo and Russell had a downright catastrophic effect when it became known throughout the world of mathematics. Confronted by these paradoxes, Dedekind and Frege completely abandoned their point of view and retreated. Dedekind hesitated a long time before permitting a new edition of his epoch-making treatise *Was sind und was sollen die Zahlen* to be published. In an epilogue, Frege too had to acknowledge that the direction of his book *Grundgesetze der Arithmetik* was wrong. Cantor's doctrine, too, was attacked on all sides. So violent was this reaction that even the most ordinary and fruitful concepts and the simplest and most important deductive methods of mathematics were threatened and their employment was on the verge of being declared illicit. The old order had its defenders, of course. Their defensive tactics, however, were

too fainthearted and they never formed a united front at the vital spots. Too many different remedies for the paradoxes were offered, and the methods proposed to clarify them were too variegated.

Admittedly, the present state of affairs where we run up against the paradoxes is intolerable. Just think, the definitions and deductive methods which everyone learns, teaches, and uses in mathematics, the paragon of truth and certitude, lead to absurdities! If mathematical thinking is defective, where are we to find truth and certitude?

There is, however, a completely satisfactory way of avoiding the paradoxes without betraying our science. The desires and attitudes which help us find this way and show us what direction to take are these:

1. Wherever there is any hope of salvage, we will carefully investigate fruitful definitions and deductive methods. We will nurse them, strengthen them, and make them useful. No one shall drive us out of the paradise which Cantor has created for us.
2. We must establish throughout mathematics the same certitude for our deductions as exists in ordinary elementary number theory, which no one doubts and where contradictions and paradoxes arise only through our own carelessness.

Obviously these goals can be attained only after we have fully elucidated *the nature of the infinite*.

We have already seen that the infinite is nowhere to be found in reality, no matter what experiences, observations, and knowledge are appealed to. Can thought about things be so much different from things? Can thinking processes be so unlike the actual processes of things? In short, can thought be so far removed from reality? Rather is it not clear that, when we think that we have encountered the infinite in some real sense, we have merely been seduced into thinking so by the fact that we often encounter extremely large and extremely small dimensions in reality?

Does material logical deduction somehow deceive us or leave us in the lurch when we apply it to real things or events?[1] No! Material logical deduction is indispensable. 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 recognizing that there are such preconditions that must be taken into account, we find ourselves in agreement with the philoso-

[1][Throughout this paper the German word 'inhaltlich' has been translated by the words 'material' or 'materially' which are reserved for that purpose and which are used to refer to matter in the sense of the traditional distinction between matter or content and logical form. – Tr.]

phers, 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 reduction. 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.

Consider the nature and methods of ordinary finitary number theory. It can certainly be constructed from numerical structures through intuitive material considerations. But mathematics surely does not consist solely of numerical equations and surely cannot be reduced to them alone. Still one could argue that mathematics is an apparatus which, when applied to integers, always yields correct numerical equations. But in that event we still need to investigate the structure of this apparatus thoroughly enough to make sure that it in fact always yields correct equations. To carry out such an investigation, we have available only the same concrete material finitary methods as were used to derive numerical equations in the construction of number theory. This scientific requirement can in fact be met, i.e., it is possible to obtain in a purely intuitive and finitary way – the way we attain the truths of number theory – the insights which guarantee the validity of the mathematical apparatus.

Let us consider number theory more closely. In number theory we have the numerical symbols

$$1, 11, 111, 11111$$

where each numerical symbol is intuitively recognizable by the fact it contains only 1's. These numerical symbols which are themselves our subject matter have no significance in themselves. But we require in addition to these symbols, even in elementary number theory, other symbols which have meaning and which serve to facilitate communication; for example the symbol 2 is used as an abbreviation for the numerical symbol 11, and the numerical symbol 3 as an abbreviation for the numerical

symbol 111. Moreover, we use symbols like $+$, $=$, and $>$ to communicate statements. $2+3=3+2$ is intended to communicate the fact that $2+3$ and $3+2$, when abbreviations are taken into account, are the self-same numerical symbol, viz., the numerical symbol 11111. Similarly $3>2$ serves to communicate the fact that the symbol 3, i.e., 111, is longer than the symbol 2, i.e., 11; or, in other words, that the latter symbol is a proper part of the former.

We also use the letters $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ for communication. Thus $\mathfrak{b}>\mathfrak{a}$ communicates the fact that the numerical symbol $\mathfrak{b}$ is longer than the numerical symbol $\mathfrak{a}$. From this point of view, $\mathfrak{a}+\mathfrak{b}=\mathfrak{b}+\mathfrak{a}$ communicates only the fact that the numerical symbol $\mathfrak{a}+\mathfrak{b}$ is the same as $\mathfrak{b}+\mathfrak{a}$. The content of this communication can also be proved through material deduction. Indeed, this kind of intuitive material treatment can take us quite far.

But let me give you an example where this intuitive method is outstripped. The largest known prime number is (39 digits)

$$\mathfrak{p}=170\,141\,183\,460\,469\,231\,731\,687\,303\,715\,884\,105\,727$$

By a well-known method due to Euclid we can give a proof, one which remains entirely within our finitary framework, of the statement that between $\mathfrak{p}+1$ and $\mathfrak{p}!+1$ there exists at least one new prime number. The statement itself conforms perfectly to our finitary approach, for the expression 'there exists' serves only to abbreviate the expression: it is certain that $\mathfrak{p}+1$ or $\mathfrak{p}+2$ or $\mathfrak{p}+3\ldots$ or $\mathfrak{p}!+1$ is a prime number. Furthermore, since it obviously comes down to the same thing to say: there exists a prime number which is

1. $>\mathfrak{p}$, and at the same time is
2. $\leqslant \mathfrak{p}!+1$,

we are led to formulate a theorem which expresses only a part of what the euclidean theorem expresses; viz., the theorem that there exists a prime number $>\mathfrak{p}$. Although this theorem is a much weaker statement in terms of content – it asserts only part of what the euclidean theorem asserts – and although the passage from the euclidean theorem to this one seem quite harmless, that passage nonetheless involves a leap into the transfinite when the partial statement is taken out of context and regarded as an independent statement.

How can this be? Because we have an existential statement, 'there exists'! True, we had a similar expression in the euclidean theorem, but there the 'there exists' was, as I already mentioned, an abbreviation for: either $\mathfrak{p}+1$ or $\mathfrak{p}+2$ or $\mathfrak{p}+3\ldots$ or $\mathfrak{p}!+1$ is a prime number – just as when, instead of saying 'either this piece of chalk or this piece or this piece … or this piece is red' we say briefly 'there exists a red piece of chalk among these pieces'. A statement such as 'there exists' an object with a certain property in a finite totality conforms perfectly to our finitary approach. But a statement like 'either $\mathfrak{p}+1$ or $\mathfrak{p}+2$ or $\mathfrak{p}+3$ . . . or (ad infinitum) . . . has a certain property' is itself an infinite logical product. Such an extension into the infinite is, unless further explanation and precautions are forthcoming, no more permissible than the extension from finite to infinite products in calculus. Such extensions, accordingly, usually lapse into meaninglessness.

From our finitary point of view, an existential statement of the form 'there exists a number with a certain property' has in general only the significance of a partial statement; i.e., it is regarded as part of a more determinate statement. The more precise formulation may, however, be unnecessary for many purposes.

In analyzing an existential statement whose content cannot be expressed by a finite disjunction, we encounter the infinite. Similarly, by negating a general statement, i.e., one which refers to arbitrary numerical symbols, we obtain a transfinite statement. For example, the statement that if $\mathfrak{a}$ is a numerical symbol, then $\mathfrak{a}+1=1+\mathfrak{a}$ is universally true, is from our finitary perspective *incapable of negation*. We will see this better if we consider that this statement cannot be interpreted as a conjunction of infinitely many numerical equations by means of 'and' but only as a hypothetical judgment which asserts something for the case when a numerical symbol is given.

From our finitary viewpoint, therefore, we cannot argue that an equation like the one just given, where an arbitrary numerical symbol occurs, either holds for every symbol or is disproved by a counter example. Such an argument, being an application of the law of excluded middle, rests on the presupposition that the statement of the universal validity of such an equation is capable of negation.

At any rate, we note the following: if we remain within the domain of finitary statements, as indeed we must, we have as a rule very complicated logical laws. Their complexity becomes unmanageable when the expressions 'all' and 'there exists' are combined and when they occur in expressions nested within other expressions. In short, the logical laws which Aristotle taught and which men have used ever since they began to think do not hold. We could, of course, develop logical laws which do hold for the domain of finitary statements. But it would do us no good to develop such a logic, for we do not want to give up the use of the simple laws of Aristotelian logic. Furthermore, no one, though he speak with the tongues of angels, could keep people from negating general statements, or from forming partial judgments, or from using *tertium non datur*. What, then, are we to do?

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 paper. Just as $i=\sqrt{-1}$ was introduced to preserve in 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*. It is quite ironic that the deductive methods which Kronecker so vehemently attacked are the exact counterpart of what Kronecker himself admired so enthusiastically in Kummer's work on number theory which Kronecker extolled as the highest achievement of mathematics.

How do we obtain *ideal statements?* It is remarkable as well as a favorable and promising fact that to obtain ideal statements, we need only continue in a natural and obvious fashion the development which the theory of the foundations of mathematics has already undergone. Indeed, we should realize that even elementary mathematics goes beyond the standpoint of intuitive number theory. Intuitive, material number theory, as we have been construing it, does not include the method of algebraic computation with letters. Formulas were always used exclusively for communication in intuitive number theory. The letters stood for numerical symbols and an equation communicated the fact that the two symbols coincided. In algebra, on the other hand, we regard expressions containing letters as independent structures which formalize the material theorems of number theory. In place of statements about numerical symbols, we have formulas which are themselves the concrete objects of intuitive study. In place of number-theoretic material proof, we have the derivation of a formula from another formula according to determinate rules.

Hence, as we see even in algebra, a proliferation of finitary objects takes place. Up to now the only objects were numerical symbols like $1, 11, \ldots, 11111$. These alone were the objects of material treatment. But mathematical practice goes further, even in algebra. Indeed, even when from our finitary viewpoint a formula is valid with respect to what it signifies as, for example, the theorem that always

$$\mathfrak{a}+\mathfrak{b}=\mathfrak{b}+\mathfrak{a},$$

where $\mathfrak{a}$ and $\mathfrak{b}$ stand for particular numerical symbols, nevertheless we prefer not to use this form of communication but to replace it instead by the formula

$$a+b=b+a.$$

This latter formula is in no wise an immediate communication of something signified but is rather a certain formal structure whose relation to the old finitary statements,

$$2+3=3+2,$$
$$5+7=7+5,$$

consists in the fact that, when a and b are replaced in the formula by the numerical symbols 2, 3, 5, 7, the individual finitary statements are thereby obtained, i.e., by a proof procedure, albeit a very simple one. We therefore conclude that a, b, $=$, $+$, as well as the whole formula $a+b=b+a$ mean nothing in themselves, no more than the numerical symbols meant anything. Still we can derive from that formula other formulas to which we do ascribe meaning, viz., by interpreting them as communications of finitary statements. Generalizing this conclusion, we conceive mathematics to be a stock of two kinds of formulas: first, those to which the meaningful communications of finitary statements correspond; and, secondly, other formulas which signify nothing and which are the *ideal structures of our theory*.

Now what was our goal? In mathematics, on the one hand, we found finitary statements which contained only numerical symbols, for example,

$$3>2,\ 2+3=3+2,\ 2=3,\ 1\neq 1$$

which from our finitary standpoint are immediately intuitable and understandable without recourse to anything else. These statements can be negated, truly or falsely. One can apply Aristotelian logic unrestrictedly to them without taking special precautions. The principle of non-contradiction holds for them; i.e., the negation of one of these statements and the statement itself cannot both be true. *Tertium non datur* holds for them; i.e., either a statement or its negation is true. To say that a statement is false is equivalent to saying that its negation is true. On the other hand, in addition to these elementary statements which present no problems, we also found more problematic finitary statements; e.g., we found finitary statements that could not be split up into partial statements. Finally, we introduced ideal statements in order that the ordinary laws of logic would hold universally. But since these ideal statements, viz., the formulas, do not mean anything insofar as they do not express finitary statements, logical operations cannot be materially applied to them as they can be to finitary statements. It is, therefore, necessary to formalize

the logical operations and the mathematical proofs themselves. This formalization necessitates translating logical relations into formulas. Hence, in addition to mathematical symbols, we must also introduce logical symbols such as

$$\underset{\text{(and)}}{\&}\ ,\quad \underset{\text{(or)}}{\vee}\ ,\quad \underset{\text{(implies)}}{\rightarrow}\ ,\quad \underset{\text{(not)}}{\sim}$$ [2]

and in addition to the mathematical variables $a, b, c, \ldots$ we must also employ logical variables, viz., the propositional variables $A, B, C, \ldots$.

How can this be done? Fortunately that same preestablished harmony which we have so often observed operative in the history of the development of science, the same preestablished harmony which aided Einstein by giving him the general invariant calculus already fully developed for his gravitational theory, comes also to our aid: we find the logical calculus already worked out in advance. To be sure, the logical calculus was originally developed from an altogether different point of view. The symbols of the logical calculus originally were introduced only in order to communicate. Still it is consistent with our finitary viewpoint to deny any meaning to logical symbols, just as we denied meaning to mathematical symbols, and to declare that the formulas of the logical calculus are ideal statements which mean nothing in themselves. We possess in the logical calculus a symbolic language which can transform mathematical statements into formulas and express logical deduction by means of formal procedures. In exact analogy to the transition from material number theory to formal algebra, we now treat the signs and operation symbols of the logical calculus in abstraction from their meaning. Thus we finally obtain, instead of material mathematical knowledge which is communicated in ordinary language, just a set of formulas containing mathematical and logical symbols which are generated successively, according to determinate rules. Certain of the formulas correspond to mathematical axioms. The rules whereby the formulas are derived from one another correspond to material deduction. Material deduction is thus replaced by a formal procedure governed by rules. The rigorous transition from a naïve to a formal treatment is effected, therefore, both for the axioms (which, though originally viewed naïvely as basic truths, have been long treated in modern axiomatics as mere relations between concepts) and for the logical calculus (which originally was supposed to be merely a different language).

We will now explain briefly how *mathematical proofs* are formalized.

[2][Although Hilbert's original paper used '$-$' as the sign for negation, we have substituted '$\sim$' for greater conformity with the notation used in other papers in this collection. – Eds.]

I have already said that certain formulas which serve as building blocks for the formal structure of mathematics are called "axioms." A mathematical proof is a figure which as such must be accessible to our intuition. It consists of deductions made according to the deduction schema

$$\frac{\begin{array}{c}\mathfrak{S}\\ \mathfrak{S}\rightarrow\mathfrak{T}\end{array}}{\mathfrak{T}}$$

where each premise, i.e., the formulas $\mathfrak{S}$ and $\mathfrak{S}\rightarrow\mathfrak{T}$, either is an axiom, or results from an axiom by substitution, or is the last formula of a previous deduction, or results from such a formula by substitution. A formula is said to be provable if it is the last formula of a proof.

Our program itself guides *the choice of axioms for our theory of proof.* Notwithstanding a certain amount of arbitrariness in the choice of axioms, as in geometry certain groups of axioms are qualitatively distinguishable. Here are some examples taken from each of these groups:

I. Axioms for implication
  (i) $A \rightarrow (B \rightarrow A)$
  (addition of a hypothesis)
  (ii) $(B \rightarrow C) \rightarrow \{(A \rightarrow B) \rightarrow (A \rightarrow C)\}$
  (elimination of a statement)

II. Axioms for negation
  (i) $\{A \rightarrow (B \,\&\, \sim B)\} \rightarrow \sim A$
  (law of contradiction)
  (ii) $\sim\sim A \rightarrow A$
  (law of double negation)

The axioms in groups I and II are simply the axioms of the propositional calculus.

III. Transfinite axioms
  (i) $(a)A(a) \rightarrow A(b)$
  (inference from the universal to the particular; Aristotelian axiom);
  (ii) $\sim (a)A(a) \rightarrow (\exists a) \sim A(a)$
  (if a predicate does not apply universally, then there is a counterexample);
  (iii) $\sim (\exists a)A(a) \rightarrow (a) \sim A(a)$
  (if there are no instances of a proposition, then the proposition is false for all a).

At this point we discover the very remarkable fact that these transfinite axioms can be derived from a single axiom which contains the gist of the

so-called axiom of choice, the most disputed axiom in the literature of mathematics:

$$\text{(i}'\text{)} \quad A(a) \rightarrow A(\epsilon A)$$

where ϵ is the transfinite, logical choice-function.

Then the following specifically mathematical axioms are added to those just given:

IV. Axioms for identity
  (i) $a = a$
  (ii) $a = b \rightarrow \{A(a) \rightarrow A(b)\}$,

and finally

V. Axioms for number
  (i) $a + 1 \neq 0$
  (ii) The axiom of complete induction.

Thus we are now in a position to carry out our theory of proof and to construct the system of provable formulas, i.e., mathematics. But in our general joy over this achievement and in our particular joy over finding that indispensable tool, the logical calculus, already developed without any effort on our part, we must not forget the essential condition of our work. 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.

The problem of consistency is easily handled in the present circumstances. It reduces obviously to proving that from our axioms and according to the rules we laid down we cannot get '$1 \neq 1$' as the last formula of a proof, or, in other words, that '$1 \neq 1$' is not a provable formula. This task belongs just as much to the domain of intuitive treatment as does, for example, the task of finding a proof of the irrationality of $\sqrt{2}$ in materially constructed number theory – i.e., a proof that it is impossible to find two numerical symbols $\mathfrak{a}$ and $\mathfrak{b}$ which stand in the relation $\mathfrak{a}^2 = 2\mathfrak{b}^2$, or in other words, that one cannot produce two numerical symbols with a certain property. Similarly, it is incumbent on us to show that one cannot produce a certain kind of proof. A formalized proof, like a numerical symbol, is a concrete and visible object. We can describe it completely. Further, the requisite property of the last formula; viz., that it read '$1 \neq 1$', is a concretely ascertainable property of the proof. And since we can, as a matter of fact, prove that it is impossible to get a proof which has that formula as its last formula, we thereby justify our introduction of ideal statement.

It is also a pleasant surprise to discover that, at the very same time, we have resolved a problem which has plagued mathematicians for a long time, viz., the problem of proving the consistency of the axioms of arithmetic. For, wherever the axiomatic method is used, the problem of proving consistency arises. Surely in choosing, understanding, and using rules and axioms we do not want to rely solely on blind faith. In geometry and physical theory, proof of consistency is effected by reducing their consistency to that of the axioms of arithmetic. But obviously we cannot use this method to prove the consistency of arithmetic itself. Since our theory of proof, based on the method of ideal elements, enables us to take this last important step, it forms the necessary keystone of the doctrinal arch of axiomatics. What we have twice experienced, once with the paradoxes of the infinitesimal calculus and once with the paradoxes of set theory, will not be experienced a third time, nor ever again.

The theory of proof which we have here sketched not only is capable of providing a solid basis for the foundations of mathematics but also, I believe, supplies a general method for treatment fundamental mathematical questions which mathematicians heretofore have been unable to handle.

In a sense, mathematics has become a court of arbitration, a supreme tribunal to decide fundamental questions – on a concrete basis on which everyone can agree and where every statement can be controlled.

The assertions of the new so-called "intuitionism" – modest though they may be – must in my opinion first receive their certificate of validity from this tribunal.

An example of the kind of fundamental questions which can be so handled is the thesis that every mathematical problem is solvable. We are all convinced that it really is so. In fact one of the principal attractions of tackling a mathematical problem is that we always hear this cry within us: There is the problem, find the answer; you can find it just by thinking, for there is no *ignorabimus* in mathematics. Now my theory of proof cannot supply a general method for solving every mathematical problem – there just is no such method. Still the proof (that the assumption that every mathematical problem is solvable is a consistent assumption) falls completely within the scope of our theory.

I will now play my last trump. The acid test of a new theory is its ability to solve problems which, though known for a long time, the theory was not expressly designed to solve. The maxim "By their fruits ye shall know them" applies also to theories. When Cantor discovered his first

transfinite numbers, the so-called numbers of the second number class, the question immediately arose, as I already mentioned, whether this transfinite method of counting enables one to count sets known from elsewhere which are not countable in the ordinary sense. The points of an interval figured prominently as such a set. This question – whether the points of an interval, i.e., the real numbers, can be counted by means of the numbers of the table given previously – is the famous continuum problem which Cantor posed but failed to solve. Though some mathematicians have thought that they could dispose of this problem by denying its existence, the following remarks show how wrong they were: The continuum problem is set off from other problems by its uniqueness and inner beauty. Further, it offers the advantage over other problems of combining these two qualities: on the one hand, new methods are required for its solution since the old methods fail to solve it; on the other hand, its solution itself is of the greatest importance because of the results to be obtained.

The theory which I have developed provides a solution of the continuum problem. The proof that every mathematical problem is solvable constitutes the first and most important step toward its solution. . . .[3]

In summary, let us return to our main theme and draw some conclusions from all our thinking about the infinite. Our principal result is that the infinite is nowhere to be found in reality. It neither exists in nature nor provides a legitimate basis for rational thought – a remarkable harmony between being and thought. In contrast to the earlier efforts of Frege and Dedekind, we are convinced that certain intuitive concepts and insights are necessary conditions of scientific knowledge, and logic alone is not sufficient. Operating with the infinite can be made certain only by the finitary.

The role that remains for the infinite to play is solely that of an idea – if one means by an idea, in Kant's terminology, a concept of reason which transcends all experience and which completes the concrete as a totality – that of an idea which we may unhesitatingly trust within the framework erected by our theory.

Lastly, I wish to thank P. Bernays for his intelligent collaboration and valuable help, both technical and editorial, especially with the proof of the continuum theorem.

[3][At this point, Hilbert sketched an attempted solution of the continuum problem. The attempt was, although not devoid of interest, never carried out. We omit it here. – Eds.]

Remarks on the definition and nature of mathematics

HASKELL B. CURRY

This paper is a discussion, written as a result of a request of Professor Gonseth, of certain points concerning the philosophy of mathematics. It is a revision of my previous discourse, on this subject, which I now regard as inadequate. The argument is based directly on my contact with mathematics without benefit of any technical acquaintance with philosophy. I have not attempted to confine myself with what is novel; but the paper is intended to be self-contained.

The principal thesis is that mathematics may be conceived as an objective science which is independent of any except the most rudimentary philosophical assumptions. It is a body of propositions dealing with a certain subject matter; and these propositions are true insofar as they correspond with the facts. The position taken is a species of formalism, which may be called empirical formalism.

The problem of mathematical truth

There are three principal types of opinion as to the subject matter of mathematics, viz. realism, idealism, and formalism. We shall consider here the realist and intuitionist views, leaving formalism for the next section.

According to realism, mathematical propositions express the most general properties of our physical environment. Although this is the primitive view of mathematics, yet, on account of the essential role played by infinity in mathematics, it is untenable to-day.

On the idealistic view mathematics deals with the properties of mental objects of some sort. There are various varieties of this view according to the nature of these mental objects. The extremes are Platonism, which ascribes a reality to all the infinistic constructions of classical mathematics, and intuitionism, which depends on an *a priori* intuition of temporal sucession. All forms of idealism are subject to the same fundamental criticism: in the first place they are vague, and, in the second

Reprinted with the kind permission of the author and editor from *Dialectica*, 8 (1954), 228–33. The author has indicated to us that, although this paper appeared in 1954, it was written in 1939 and represents his views as of that time.

place they depend on metaphysical assumptions from which mathematics, if it is to have the pre-philosophical character above mentioned, must be free.

It is important to see that this criticism, so obvious in the case of Platonism, applies also to intuitionism. As to the vagueness, Heyting, in his Ergebnisse report, explicitly denies the possibility of an exact description of this mathematical intuition. As to the metaphysical character, it is clear from the intuitionist writings that their "ur-intuition" has the following properties: (1) it is essentially a thinking activity; (2) it is *a priori;* (3) it is independent of language; and (4) it is objective in the sense that it is the same in all thinking beings. The existence of an intuition – temporal or not – satisfying these four conditions is an outright assumption.

The formalist definition of mathematics

According to formalism the central concept in mathematics is that of a formal system. Such a system is defined by a set of conventions, which I shall call its *primitive frame,* specifying the following: first, what the objects of the theory, which I shall call *terms,* shall be; second, how certain propositions, which I shall call *elementary propositions,* may be stated concerning these terms, i.e. what *predicates* (classes, relations, etc.) we shall take as fundamental; and third, which of these elementary propositions are true. The first and third of these sets of conventions are essentially recursive definitions; we do not specify the ultimate nature of the terms, but give simply a list of primitive terms, or tokens, together with operators and rules of formation by means of which all further terms are constructed; likewise we start with a list of elementary propositions, called *axioms,* which are true by definition, and then give *rules of procedure* by means of which further elementary theorems are derived. The proof of an elementary proposition then consists simply in showing that it satisfies the recursive definition of elementary theorem.

It should be noted that in such a formal system it is immaterial what we take for the tokens (and operators) – we may take these as discrete objects, symbols, abstract concepts, variables, or what not. Any such way of understanding a formal system we may call a representation of it. The primitive frame specifies, independently of the representation, which elementary propositions are true, and therefore determines the meaning of the fundamental predicates. In this sense the primitive frame defines the system.

One representative of particular importance is when the tokens are taken as symbols. In this representation, which is insisted on by Frege and his followers (Hilbert, Carnap, the Poles), a formal system becomes essentially equivalent to the formalized syntax of an object language. This representation has certain advantages of definiteness and concreteness. But it also has certain disadvantages. For it is necessary, as Carnap has shown, to distinguish between the symbols used as names of the terms and the specimens of those terms; usually this means that the familiar symbols are used for the latter purpose and more or less outlandish ones for the former. Since we never use the symbols of the object languages in the theory but only in the introduction, this leads to unintelligibility. Why not abolish the object language altogether and understand that the tokens are objects which we can take as symbols if we want to? Again the consideration of other sorts of representations may suggest simplifications which a syntactical representation would not. Thus the syntactical viewpoint gives rise to an extreme nominalism - as shown by the inclusion of parentheses, commas, etc., among the symbols and of expressions which are not "well formed" - which is avoidable and contrary to the spirit of mathematics.

In the study of formal systems we do not confine ourselves to the derivation of elementary propositions step by step. Rather we take the system, defined by its primitive frame, as datum, and then study it by any means at our command. In so doing we may formulate further propositions, which we call metatheoretic propositions. Like the elementary propositions these state, essentially, properties of the system; but they may also involve extraneous considerations. Insomuch as they deal with what is, in view of the primitive frame, a well defined subject matter, the question of their truth involves no difficulties beyond those inherent in compound propositions in general.

The formalist definition of mathematics is then this: mathematics is the science of formal systems. The propositions of mathematics are the propositions, elementary or metatheoretic, of some formal system or set of systems. For each such proposition which does not involve extraneous considerations, we have an objective criterion of truth in the sense that an alleged proof can be checked objectively; but a proposition may be indefinite in the sense that we have no resolution process (*Entscheidungsverfahren*). Intuition is, of course, involved in this; viz. the intuition of recursive definitions, mathematical induction and the like; but the physical nature of this intuition is irrelevant. (If extraneous considerations are involved, then we have to do with applied, not pure, mathematics - the boundary line between mathematics and other sciences is not sharp, and should not be.) It should be noted that we have not confined mathematics to a single formal system; moreover, metatheoretic propositions are included in mathematics. This answers the objections which

might be raised on the ground of the incompleteness theorems of Skolem, Gödel, et al.

Truth and acceptability

We now turn to the relation of mathematics to its application. For this purpose we introduce another kind of quasi-truth concept which applies, not to single propositions, but to systems as a whole. I shall call this *acceptability*. By acceptability, then, I mean the considerations which lead us to be interested in one formal system rather than in another.

Acceptability is usually a matter of interpreting the theory in relation to some subject matter. Such an interpretation is to be distinguished from a representation: in a representation the predicates are defined by the primitive frame; in an interpretation we associate them with certain intuitive notions, so that the question arises as to the agreement between the truth of the propositions of the formal system and that of the associated intuitive ones. Acceptability is thus relative to a purpose; and a discussion of acceptability is pointless unless the purpose is stated.

As an illustration of acceptability questions let us consider the acceptability of classical mathematics for the purpose of application in physics.

Among the criteria of acceptability we may mention the following: (1) the intuitive evidence of the premises; (2) consistency (an internal criterion); (3) the usefulness of the theory as a whole.

The intuitionists have made much of the first criterion. They point out that certain propositions of classical mathematics lack intuitive evidence; and they have constructed systems which - we can admit this without swallowing the intuitionistic metaphysics - have greater intuitive evidence than the classical. But these systems are so complicated as to be useless, and are inacceptable by criterion (3). Moreover this is the decisive consideration; for physics is an empirical science, and therefore the question of intuitive evidence is secondary. The acceptability of classical mathematics is an empirical fact, and the proper retort to the intuitionist gibe, that classical mathematics has only a heuristic value, is that so far as physics is concerned, that is all the value that an intuitionist mathematics has either.

The criterion of consistency has been stressed by Hilbert. Presumably the reason for this is that he, like the intuitionists, seeks an *a priori* justification. But aside from the fact that for physics the question of an *a priori* justification is irrelevant, I maintain that a proof of consistency is neither a necessary nor a sufficient condition for acceptability. It is obviously not sufficient. As to necessity, so long as no inconsistency is known, a consistency proof, although it adds to our knowledge about the system, does not alter its usefulness. Even if an inconsistency is discovered this does not mean complete abandonment of the theory, but its modification and refinement. As a matter of fact, essentially this has happened in the past; we now know, for example, that the mathematics of the eighteenth century was inconsistent, but we have not abandoned the results of the eighteenth-century mathematicians. The peculiar position of Hilbert in regard to consistency is thus no part of the formalist conception of mathematics, and it is therefore unfortunate that many persons identify formalism with what should be called Hilbertism.

Let us now cut short this discussion and summarize as follows: Acceptability is relative to a purpose, and a system acceptable for one purpose may not be for another. For example, I agree with Weyl and Gentzen that there are purposes for which intuitionistic systems are acceptable, although they are not acceptable, on empirical grounds, for application to physics. Again, acceptability is a different question from truth; in fact a formalist definition of mathematical truth is compatible with almost any position in regard to acceptability. In this sense formalist mathematics is compatible with various philosophical views; it is an objective science which can form part of the data of philosophy.

Mathematics and logic

In current popular discussions it is said that intuitionism, formalism, and logicism are the three main views in regard to the nature of mathematics; the last is supposed to be the view that mathematics is logic. But we do not have here a third view of mathematics parallel with the other two; for to say that mathematics is logic is merely to replace one undefined term by another. When we go back of the word "logic" to its definition in the logistic systems, we find that they run the gamut from extreme Platonism to pure formalism. The question of the relation of mathematics to logic is thus a different question from the definition of mathematics; on account of the lack of space we cannot go into that question here.

Hilbert's programme

GEORG KREISEL

1. If one may judge from his publications, Hilbert's conception of the problem of foundations underwent marked developments.

In [1932-5, 3: 145-56; orig. 1918] he still concentrated on the "sound" and rather colourless *Independence Problem* which may be formulated as follows: given a branch of knowledge which is so well-developed as to be axiomatized, the problem is to get a clear view of the logical relationships (dependence and independence or, derivability and non-derivability) of statements of the axiomatic theory. Hilbert emphasized the consistency problem which is so to speak the weakest non-derivability result, since it is the problem of showing that there exists at least one statement which is not derivable.

But in later writings (though also in 1905, [1899b, 7th ed.: 247-61] the *Consistency Problem* was associated with the problem of understanding the concept of infinity. He sought such an understanding in understanding the *use of transfinite machinery* from a finitist point of view. And this he saw in the elimination of transfinite (ϵ—) symbols from proofs of formulae not containing such symbols. He was convinced from the start that such an elimination was possible, and expressed it by saying that the problems of foundations were to be *removed* or that doubts were to be eliminated instead of saying that they were to be investigated.

We note at once that there is no evidence in Hilbert's writings of the kind of formalist view suggested by Brouwer when he called Hilbert's approach "formalism." In particular, we could say that Hilbert wanted to eliminate the use of transfinite concepts from proofs of finitist assertions instead of referring to symbols and formulae as above. The symbols were a means of representation. The real opposition between Brouwer's and Hilbert's approach was not at all between formalism and intuitive mathematics, but between (i) the conception of what constitutes

Revised by the author from an earlier version which appeared in *Dialectica* 12 (1958), 346-72. Printed here by kind permission of the author and the editor of *Dialectica*. For this second edition Professor Kreisel has added a Postscript, as well as notes to a number of the sections of the original article. These are collected at the end of the essay. Except as noted otherwise, these additions were made in the autumn of 1978 and represent his views as of that time.

a foundation[1] and (ii) between two informal ways of reasoning, namely finitist and intuitionist. In fact, Bernays repeatedly emphasized the latter point, the lack of evidence in the basic intuitionistic conception of constructive proof [Hilbert 1932–5, 3: 212, or Bernays 1941: 147]: in short, it is not the restrictions imposed *by* intuitionism, but those *on* intuitionism which seem to constitute the most significant differences. Hilbert's own remarks on this opposition seem quite inept.[2]

The view above on the significant differences between Brouwer's and Hilbert's approach does not deny, of course, the popular attraction of a syntactic formulation of foundational problems. Derivability in formal systems (which codify the manipulation of symbolic representations of transfinite concepts) is "down to earth," it refers to palpable acts, to what we "actually do," while the transfinite concepts themselves are "up in the air," they are abstract, and therefore supposed to be inaccessible to exact study. Even though Hilbert was not a strict positivist like Comte [Hilbert 1932–5, 3: 387], his presentation was certainly congenial to a positivist era: at least, his problems were positivistically meaningful for a more liberal conception of positivism, while the transfinite concepts themselves are senseless even for the latter. It might have been expected that deficiencies in our understanding of these concepts would reappear in our inability to solve the syntactic problems, in particular to survey those parts of the corresponding formal system which are devoid of positivist interest, namely all those formulae which contain symbols for transfinite concepts. But equally it was to be expected that specific partial (syntactic) problems could be solved despite a lack of deeper understanding of the abstract concepts. In the following sections we shall try to "reconstruct" Hilbert's programme in the light of these observations.

Hilbert was certainly not a fanatic of crude formalism: thus while his problems concerned syntactic properties of formal systems, the solutions were to be given by intuitively correct reasoning, and he explicitly considered any formalization of *this* reasoning as unnecessary. (For qualifications, cf. Section 18.)

2. The fabric of Hilbert's conception. He asserted that there was a certain type of evident reasoning which was presupposed in all scientific thinking [Hilbert 1932–5, 3: 162–63], and finitist operations were typical

[1]Brouwer ignores non-constructive mathematics altogether and therefore does not have an analogous problem of foundations to Hilbert's.

[2]E.g., [Hilbert 1899b, 7th ed.: 307]: Considering that the intended meaning of the intuitionistic disjunction is different from that of classical disjunction, the rejection of *tertium non datur* is much more like depriving non-commutative algebra of the rule $ab=ba$ than a boxer of the use of his fists.

of this. He believed that there were no essentially different truths in mathematics [Hilbert 1932–5, 3: 157].

These views lead to the hope of a final solution of the problem of foundations [Hilbert 1931: 489, 494] by a reduction of all mathematical reasoning to finitist reasoning: for if the minimum that has to be presupposed suffices for this reduction then we have a complete solution. Conversely, something of this kind is required for a "complete" solution: if there is a plurality of essentially different mathematical truths there is hardly a hope of an enumeration of such truths which convinces us as being complete; furthermore there would be the problem of their interrelations, and so there would be no unique outstanding problem of foundations.

Next, if the finitist truths are the only[3] absolute ones [Hilbert 1932–5, 3: 180], it is at least natural to regard mathematical expressions containing transfinite symbols as "ideal" elements whose sole purpose is the streamlining of the symbolism [Hilbert 1899b, 7th ed.: 280, or 1932–5, 3: 187]. And in this case consistency is all that matters because, in the usual systems, if consistency is established by (finitist) methods, and a formula without transfinite symbols is derived in the system considered, then this formula can be proved by the same methods [Hilbert 1899b, 7th ed.: 304, and Bernays 1941: 154]; in other words, we have the required elimination. Conversely, if consistency is all that we demand then there is no assurance that formulae containing transfinite symbols in an essential way have the intended interpretation, and so they must be regarded as ideal. This is established by Gödel's construction of consistent, but ω-inconsistent, systems,[4] because purely universal formulae are deductively equivalent to formulae without transfinite symbols, and purely existential formulae which are provable in a consistent system need not be true.

We note in passing an interesting aspect of Hilbert's idea of a *paradise:* a characteristic of Cantor's set theory (Hilbert's paradise, 1899b, 7th ed.: 274) is the abundance of transfinite machinery which Hilbert regarded in the same paper as "ideal" elements to be used as gadgets to make life smoother.

[3]Though, I think, Hilbert does not deny this elsewhere, his emphasizing that the geometric continuum [1934–5, 3: 159] is a concept in its own right and independent of number seems to weaken the doctrine that all absolute truths are finitist.

[4]The possibility of ω-inconsistent systems was evidently clear to Hilbert, but as late as 1930 [1899b, 7th ed.: 320] he wanted to show that every consistent statement of arithmetic was provable, i.e., that the usual system of arithmetic was complete. Actually he requires more, namely that consistency "looks after the rest", since a system might be complete and yet some of its theorems false.

It is plain that such concepts as "finitist," "essentially the same truth," "reduction" are not at all precise. But if one really believes in the success of the finitist reduction it was not necessary to clarify them in advance. For when the work is done one can examine what methods are actually needed, in what sense we have a reduction, and at this stage one can then decide if it is satisfactory.

3. *Critique of detail.* There is one point in the above picture which is not convincing even if the basic assumptions are granted. Hilbert regarded complex numbers [1899b, 7th ed.: 269] as a typical example of ideal elements. Yet the reduction to pairs of real numbers does not only ensure consistency, but also gives each formula containing symbols for "ideal" elements a *meaning* in terms of real numbers. It would not seem unreasonable to demand the same for formulae with transfinite symbols, at least in any specific context. At any rate one would thereby extract more "absolute truths" from the formal machinery.

4. *Critique of basic assumptions.* Of course, the first basic assumption is that a reduction to finitist methods is possible at least in the sense of a finitist consistency proof. If this assumption is false, the problems at the end of para. 2 reappear and more besides.

First, instead of having a single kind of elementary reasoning whereby we understand the use of transfinite symbols, there will now be methods of reasoning involving a hierarchy of conceptions such as, e.g. more and more abstract conceptions of a "construction," and we have a hierarchy of Hilbert programmes of *discovering the appropriate complex of such methods which is needed for understanding the use of transfinite symbols in given systems* (modified Hilbert programme).

Second, it will be necessary to ascertain that the consistency of a particular system cannot be established by finitist means, or whatever other complex of methods is being considered. For such an impossibility result it will be necessary to define the notion of *finitist proof* and even to make precise how *consistency* is to be formulated. The latter point is illustrated by the need for derivability conditions on the arithmetized proof predicate in Gödel's second undecidability theorem.

Third, when we are not dealing with an elimination of the nonfinitist methods, but with a *separation* between them, it is necessary to determine the significance of the distinction between finitist and non-finitist. An analogous problem will arise for each subclass of the constructive methods used in the analysis of the transfinite machinery.

Hilbert's own writings contain little information about the solution of these new problems. For the first problem Gentzen's use of ordinals $<\epsilon_0$

is a good illustration of the kind of subclass of constructive methods which is particularly appropriate for the analysis of a given class of transfinite methods, in this case number theory.

It is difficult to separate the second and the third problem in practice because one can only decide whether e.g. a *definition* of finitist proof is correct if its *significance* (importance) is clear. Hilbert himself is quite unconvincing about the inherent virtues of finitist reasoning. At one time [1932–5, 3: 160, 162] the main purpose of the finitist reduction, in fact the whole need for foundations, consisted for him in clearing the fair name of mathematics which had been sullied by the paradoxes. Now on p. 158 of the same paper he said that the paradoxes simply have nothing to do with the theory of sets of numbers: it is hard to see why this remark, if true, has not cleared the fair name at least of analysis unless one believes that stained reputations can only be cleared with a great deal of ceremony. Von Neumann's line [1927] on the subject is that finitist consistency proofs would reduce strict intuitionism *ad absurdum* (but only if one means by strict intuitionism the view that classical analysis is formally inconsistent and not merely in contradiction with intuitionistic theorems): this is at least less pious than the talk about reputations, but doing down the intuitionists is hardly a grand scientific programme. – Hilbert did emphasize the increase of information contained in proofs of a formula of the form $(\mathrm{E}x)(y)\mathrm{A}(x,y)$ [1932–5, 3: 154, 155] from a pure existence proof [intuitionistically $\sim(x)\sim(y)\mathrm{A}(x,y)$] to $(\mathrm{E}x)[x\leqslant 12\ \&\ (y)\mathrm{A}(x,y)]$ to $\mu_x(y)\mathrm{A}(x,y)=10$. But this gives no clue to the nature of the improvement involved in replacing a non-finitist proof of a universal formula $(x)\mathrm{A}(x)$ by a finitist one, and this is the critical case (cf. Kreisel 1958b: 177).

Hilbert sometimes speaks of the reliability (*Sicherheit*) of finitist reasoning. As Bernays has pointed out [Hilbert 1932–5, 3: 210], realistically speaking, almost the opposite is true, the chance of an oversight in long finitist arguments of metamathematics being particularly great. At any rate it seems improbable that a satisfactory characterization of "finitist proof" would be based on this notion of reliability. – We shall take up the subject in the text below.

5. *Critique of basic assumptions* (continued). In the previous paragraph we considered some problems which arise when one attempts to follow Hilbert's aim of understanding the concept of infinity by eliminating the use of transfinite machinery from proofs of finitist or, more generally, constructive assertions. But this view, that understanding a concept consists simply in the technique of reasoning about it in some well-defined context, does not seem quite adequate.

Thus e.g. the first-order theory of the addition of natural numbers has a complete formalization and hence a decision method. So, from the syntactical point of view it leaves nothing to be desired. But by the compactness theorem for the predicate calculus this theory has a model containing "non-standard" integers. Thus we do not get the degree of understanding that we should have with a system which is satisfied only by a finite set. – We may regard this as a limitation on the syntactic approach for an understanding of the concept of infinity or, at least, as illustrating the use of the notion of a *model* in this connection.

Usually Gödel's incompleteness theorems are taken as showing a limitation on the syntactic approach to an understanding of the concept of infinity. We note the superficially paradoxical fact, brought out above, that the completeness of the predicate calculus, so to speak the "adequacy" of the syntactical approach for predicate logic, leads to a limitation too. This is a point of view emphasized by Skolem.

6. Conclusion. In my opinion, Hilbert's programme, including the modified version in section 4 above and the independence problem of section 1, is a rich line of research in foundations. The general problems of section 4 seem important and somewhat more specific problems will be arrived at below in a brief analysis of the work done by Hilbert's school so far. Also, the (modified) Hilbert programme gives scope to a great variety of methods of mathematical logic including those of the topological approach, intuitionism, recursion theory, as will be seen below.

My own attitude towards the original Hilbert programme is this.

As far as piecemeal understanding is concerned, its importance consists in having led to the fruitful study of the constructive aspects of axiomatic systems. But even if it is compared only with other parts of mathematical logic and not with other mathematical disciplines, its role is not unique; cf. the studies of the distinction between first-order and higher-order reasoning, or of the set theoretical aspects of informal mathematics. My own interest in the modified Hilbert programme does not go one way only, i.e. the elimination of non-constructive methods, but I find that greater facility with the non-constructive methods comes from a study of their constructive aspects.[5]

As far as an over-all philosophical understanding is concerned, the original Hilbert programme has failed, and, as is usual with great schemes it gives no hint of what might take its place. When asked: "What is mathematics about?", Hilbert could still have said: about the arithmetico-combinatorial facts of finitist mathematics; even though the latter may raise problems of their own, such a "reduction" could have been satisfying. Hilbert's answer is simply not true even for the very weak sense of "equivalence of content" expressed in statements of formal deducibility and nondeducibility. Furthermore, we have no idea by what sort of investigation we could even hope to find a satisfactory answer to such a question. Hilbert thought it would be supplied by pure mathematics itself [1899b, 7th ed.: 316]. But it seems clear, as Bernays has expressed it, that the totality of pure mathematics (mathematical structures) is not itself a mathematical structure; this is not only a stumbling block to a mathematical treatment of the conception of the whole of mathematics, but even to an exhaustive treatment of the concept of natural or real number because the characterization by means of a least[6] (or largest) class refers explicitly to the totality from which these classes are to be taken. If one may use for orientation the formulation of *finitist proof* sketched below, one would say this: just as it is necessary to use non-finitist concepts to study the totality of finitist proofs, so it is necessary to use non-mathematical concepts, i.e. concepts lacking the precision which permit mathematical manipulation, for a significant approach to foundations. It is not at all a question of concepts which are more "reliable" than those of current mathematics, but of concepts which provide a frame of reference for discussing the status of mathematics (cf. Bernays 1957: 245). I do not believe that at the present time we have any concepts of this kind which we can take seriously.

7. For certain parts of non-constructive mathematics the Hilbert programme has been carried out in the originally intended sense, e.g. arithmetic without induction based on the classical predicate calculus. For this purpose a detailed syntactic analysis of the latter is used. Below we shall describe three methods of syntactic analysis, their mathematical

[5] It is remarkable that mathematicians have not yet learned to exploit the non-constructive methods effectively in the following sense: in the famous non-constructive proofs of constructive results (for references, see Kreisel 1958b) the elimination of the non-constructive methods used is finitist, and does not require more sophisticated notions of constructivity. We know that a more essential use of non-constructive methods must be possible, and, I believe, closer study of their constructive aspects may give one a better "feeling" for them.

[6] It is of course possible (and natural) to consider the relevant extremal causes as primitive notions in their own right and not as part of set theory, e.g., the "and only those required by the foregoing rules" in the usual definitions of natural numbers or recursive ordinals. This is used, e.g., in Hilbert's reduction [1899b, 7th ed.: 302] of the principle of induction to the reversibility of the formation of numerals. The latter follows from the extremal clause, and in a similar way other principles of proof can be obtained from such clauses (cf. Lorenzen's induction and inversion principles). But in the application of induction the troublesome totalities reappear in the choice of properties to which induction is applied, e.g., by applying induction to certain non-elementary properties such as truth-definitions formulae of classical arithmetic Z can be proved which cannot be proved in Z itself.

significance, including applications to the independence problem, and their bearing on the modified Hilbert programme. We conclude with two observations on finitist proofs and the completeness of predicate logic.

I. Syntactic analysis

8. We begin by considering what seems to be the main novelty of Hilbert's work in mathematical logic, without which his conception of the Hilbert programme would have been wholly unconvincing, namely his proof theory. He wanted *proof* itself to be made the object of mathematical study [1932–5, 3: 165]. Though this is needed for a syntactic (combinatorial) *formulation* of independence results, by itself it is not the crucial point for proof theory. For, clearly, the traditional independence proofs by means of models, as in the case of the parallel axiom, or even the impossibility proofs for certain constructions by means of ruler and compass, are applicable to formalized systems. Thus the consistency of the rules of set theory is proved as follows: when read in the intended manner, the axioms of, e.g., Zermelo's set theory are true of the concept of set and the rules of proof are such that true statements are transformed into true ones. Hence the formal system is consistent.

No, from the point of view of technique the crucial point is that from an early stage Hilbert had in mind a *new type of analysis* in which the detailed structure of the proof is considered. In particular, in the consideration of the so-called transfinite symbols $\epsilon_x A(x)$, one does not define "models" for them once and for all, but different numbers are substituted for a given symbol depending on the particular proof of the system which is analyzed. Briefly: instead of constructing a model for a system as a whole he gives a method for constructing a model for each particular proof of the system.

In short, Hilbert's conception of a mathematical theory in which *proof* itself is an object of study does not restrict the means of independence proofs but on the contrary enlarges it.[7] A restriction comes about only when one restricts the methods to be used in this new theory to so-called finitist or arithmetico-combinatorial methods. We observe in passing the superficially paradoxical character of the fact that up to now the significant independence proofs for classical (and intuitionistic) arithmetic have been obtained by the restricted and not by the potentially more powerful methods. We shall return to this point below.

[7]Within the framework of classical constructive mathematics (i.e., non-constructive methods of proof, but only recursive functions and predicates) we can give this the following precise sense: by means of (quantifier-free) double induction syntactic independence proofs for primitive recursive arithmetic can be obtained, but not by means of constructive models [e.g., Kreisel 1958b or Mostowski 1957a].

9. Decision problem. A very satisfactory syntactic analysis is got from a practical decision method for the theory considered. In particular it solves the independence problem of para. 1 for finite sets of statements of the theory considered. It is not the most satisfactory solution because, given the decision method, one can now ask whether given statements are independent of a certain *infinite* set of statements of the theory, and this question need not be decidable; in fact, it seems inconceivable that there is an optimal solution of the problem. Similarly, the impossibility of a decision method for a given theory is not a "catastrophe";[8] it sets a limit on how[9] clear a view one may reasonably expect to get of the structure of the theory. – This stoic view of the matter is probably universal now, perhaps largely due to the thoughtful writings of Bernays.

It is clear that if a system is consistent a (finitist) proof of decidability automatically yields a (finitist) consistency proof. As early as 1927, von Neumann doubted the decidability of the predicate calculus, though Hilbert does not seem to have been quite so definite. But in any case his actual investigations aimed at much less than decidability. We shall now describe them.

10. The ϵ-substitution method. The main tool in his work on predicate logic was a reformulation of the predicate calculus by means of the ϵ-symbol.

This reformulation is not specially elegant in practice, e.g. the formula $(Ey)(z)A(y,z)$ in the usual notation is written as

$$A[\epsilon_y A\{y, \epsilon_z \neg A(y,z)\}, \epsilon_z \neg A\{\epsilon_y A[y, \epsilon_u \neg A(y,u)], z\}],$$

but it makes *evident* a fact of logical reasoning which was basic for Hilbert's programme, namely that in logical reasoning one never makes full use of the intended meaning of the transfinite symbols, in the following sense. $\epsilon_z A(y,z)$ is intended as a choice function (some z_1 which satisfies $A(y,z)$ if such a z_1 exists, and an arbitrary value otherwise), and on this meaning the schema

$$(*) \qquad A(y,b) \rightarrow A[y, \epsilon_z A(y,z)]$$

is valid. Also, in an obvious way, the universal and existential quantifier can be defined by the use of the ϵ-symbol, and the schema (*) is enough to derive the usual schemata for the quantifiers. Now the gain is this:

[8]The Twenties constantly saw potential catastrophes in the doings of logicians.

[9]It is usual to measure this by the degree of undecidability of the theory. There is quite a different conception of a "partially clear view," namely that obtained by methods enumerating (subsets of the) unprovable formulae, as afforded by incomplete interpretations (cf. para. 15).

we see immediately that in any given proof, since (*) is applied only to finitely many y, one never needs the full extension of the choice function $\epsilon_z A(y,z)$ for all y, and, moreover, in the course of the proof one does not need its "real" values, but, e.g. if (*) were the only application of the schema in the given proof, we could simply take b for $\epsilon_z A(y,z)$ and still have a proof. Such considerations make the elimination of the ϵ-symbols (from a proof of a formula without ϵ-symbols) at least plausible. This idea was developed by Ackermann, presented in detail in [Hilbert and Bernays 1934–9: vol. 2], for predicate logic, and for number theory by Ackerman [1940]. – However, for practical applications it is best to apply the simple idea directly [Kreisel 1958b: 171].

There is a formulation of the substitution problem which does not use the ϵ-symbol at all. If distinct ϵ-matrices are replaced by distinct function symbols f, the ϵ-formulae reduce to the form $\Phi(f_1,\ldots,f_n)=0$ where Φ is an elementary functional. The problem is to prove $(Ef_1)\ldots(Ef_n)$ $[\Phi(f_1,\ldots,f_n)=0]$; it is evident if this is true at all there are functions f^* which are zero except for a finite number of arguments, and can therefore be found by trial and error. The existence of functions f seems assured by the interpretation of ϵ-matrices as choice functions.

It seems understandable that Hilbert assumed that the *proof* of such an elementary matter as the existence of f^* must be a relatively minor task.[10] For, if one regards metamathematical results as the absolute truths of mathematics [Hilbert 1932–5, 3: 180] and the "transfinite" formulae as ideal elements without significance outside the framework of a formal system, it is natural to regard the metamathematical results as significant independently of their proof: though this, of course, does not mean that the proof is easy or even of an elementary character, one may be tempted to think so, e.g. because of the double meaning of "significant" (meaningful, but also: not trivial, not easy).

Open problem. Notwithstanding the interest of alternative analyses of predicate logic and its extensions, to be described below, an examination of the substitution method applied to analysis seems very promising at the present time. Somewhere there is a combinatorial lemma lurking in the proofs which show that the substitution method terminates and which gives information about the solution of the functional equations $\Phi(f_1,\ldots,f_m)=0$ mentioned above.[11]

[10]In [1899b, 7th ed.: 317], Hilbert expresses this by saying that in the case of analysis "only" the proof of the purely arithmetical statement that the method terminates is needed. (He had of course the feeling [1932–5, 3: 187] that a purely arithmetical truth must have a purely arithmetical proof, which is refuted by Gödel's theorem if, e.g., "arithmetical" is interpreted as: expressible in classical number theory.) – Hilbert's "only" is misplaced: for after all, the statement of consistency is also purely arithmetical, so from the very start, "only" the proof of the main contention of his life was needed. However, from a mathematician's point of view, Hilbert's excitement at his reformulation of the consistency problem in terms of the convergence of the substitution method is very natural: the latter problem, and particularly the problem of finding bounds, has the general look of a mathematical problem, while the consistency problem does not, even if it is formulated as a combinatorial problem (cf. end of para. 10).

11. Cut free formalizations. Other syntactic analyses of predicate logic were developed by Herbrand and perfected by Gentzen. They resulted in reformulations of predicate logic specially adapted for proofs of completeness.

Herbrand's *lemme fondamental* becomes much simpler for prenex formulae[12] in contrast to Gentzen's *Hauptsatz*. Be that as it may, both of them gave explicit (primitive recursive) instructions for converting a proof with cuts into one without. But the simplest[13] exposition of, e.g. Herbrand's reformulation of the predicate calculus goes by way of the completeness theorem: if a (prenex) formula is not provable by Herbrand's rules (Hilbert and Bernays 1934–9, 2: 158, *b*) then it is not valid; so if a formula is provable in the ordinary way it must be provable by Herbrand's rules. But though further constructive analysis of the completeness proof is possible (cf. Kreisel 1958b: 168), without it the rule for getting an Herbrand proof is only general recursive.

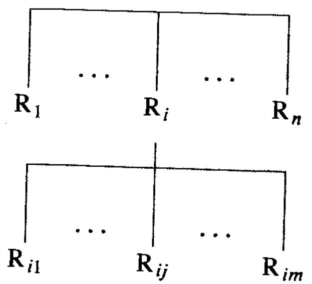
Given a prenex formula, say $(x)(Ey)(z)A(x,y,z)$, we ask: how could it be false? i.e. $(Ex)(y)(Ez)\neg A(x,y,z)$. There would have to be an element α such that $(y)(Ez)\neg A(\alpha,y,z)$ and without loss of generality we may as well call it 0, i.e. $(y)(Ez)\neg A(0,y,z)$. For $y=0$ we must have $(Ez)\neg A(0,0,z)$; we may as well take $z=1$; for either $z\neq 0$, then calling $z=1$ is permissible, if $z=0$, then we must simply regard 1 as another name for the individual called 0; we must not take $z=0$ in general. This explains the disparateness conditions of [Hilbert and Bernays 1934–9, 2: 173]. Now in order that $\neg A(0,0,1)$, there is a certain finite set of truth distributions on the prime formulae of $\neg A(0,0,1)$ which may be recorded in the form of a finitary tree. Let R_i be the conjunction of the

[11]This "lurking lemma" has since been formulated and proved by Tait [1965a] for the substitution method as applied to the formalism of arithmetic. It shows in particular in a natural way how the first ϵ-number enters into the problem: so it represents one of three independent analyses of the role of ϵ_0, the other two being the computational analysis of Gödel's functionals of para. 12 below, given in detail in [Tait 1965b], and the analysis of infinite cut free proof trees by means of ordinals (cf. footnote 15).

[12]The existence of prenex normal forms in classical logic makes this distinction more important in the analogous treatment of intuitionistic logic.

[13]I have not studied Herbrand's own publication. The exposition is suggested by Beth's semantic tableaux [1957] which are a generalization of the criteria of refutability [Hilbert and Bernays 1934–9: vol. 2]. – The proof of Herbrand's theorem [Hilbert and Bernays 1934–9: vol. 2] gives primitive recursive instructions.

ith set of prime formulae and negations of prime formulae with arguments 0, 1 which make $\neg A(0,0,1)$ true. Now, for each i we consider all extensions $R_{i1}, \ldots, R_{im}$ of R_i by prime formulae with arguments 0, 1, 2 which make $\neg A(0,1,2)$ true, $\neg A(0,0,1)$ being made true automatically since R_{ij} are extensions of R_i. We record this information in a tree



(If there is no extension of R_i, we stop the construction of R_{ij} and consider R_{i+1}, $i<n$.)

If this tree is unbounded, by the *Unendlichkeitslemma*, there is an infinite branch. If the prime formulae of $(x)(Ey)(z)A(x,y,z)$ are given the truth values which they have on such a branch and if the variables range over the natural numbers then $\neg(x)(Ey)(z)A(x,y,z)$. (Completeness.)

If the tree is bounded, the formula $\neg(x)(Ey)(z)A(x,y,z)$ is not satisfiable at all. Now the whole tree can be converted into a proof of $(x)(Ey)(z)A(x,y,z)$. For, regard the numerals as variables. All we need are rules which allow us to infer $(x)(Ey)(z)A(x,y,z)$ from $A(0,0,1) \vee A(0,1,2) \vee \ldots \vee A(0,n,n+1)$ for each n. The rules required are just Herbrand's rules. Thus we have not only a new formalization of, but also a cogent motivation for, the choice of the rules. – Beth [1957] has used the same idea for treating arbitrary formulae and arriving at a variant of Gentzen's rules.

Here one conceives of a counter example to $(x)(Ey)(z)A(x,y,z)$ in terms of satisfying $\neg A[0,n,\varphi(n)]$ for some given disparate function φ and arbitrary numerals n. The impossibility of obtaining one, i.e. the breakdown of the construction above, yields a classical proof of $(x)(Ey)(z)A(x,y,z)$. Another conception of a counter example is to satisfy $\neg(x)(Ey)(z)A(x,y,z)$ in the prenex form $(Ex)(y)(Ez)\neg A(x,y,z)$ by a constant α and a function f, i.e. $(y)\neg A[\alpha,y,f(y)]$. The impossibility of obtaining one is expressed by $(Ey)A[\alpha,y,f(y)]$ and this can be expressed explicitly by means of functionals $\varphi_i(f,\alpha)$ (terms containing α and f) such that $\ldots \vee A[\alpha,\varphi_i(f,\alpha),f\{\varphi_i(f,\alpha)\}] \vee \ldots$[14] — This is only

[14]This is used in [Hilbert and Bernays 1934–9: vol. 2] as an auxiliary in the proof of Herbrand's theorem. Below and in other publications we emphasize the independent sig-

suitable for prenex formulae (unless one uses functionals of higher type) but more suitable for arithmetic than the alternative described above.

12. Gödel's intervention. There is a totally different analysis of classical proofs due to Gödel. First classical proofs are replaced by intuitionistic ones of (classically) equivalent theorems [Gödel 1931–2b], then these proofs are analyzed by means of certain simple functionals of finite type, and these in turn are shown to be well defined by means of transfinite induction ($<\epsilon_0$ for arithmetic with induction, $<\omega^\omega$ for arithmetic without). In this way the syntactic analysis is effected in several manageable steps, each of them of interest in itself.[15] For further details see Gödel's article in *Dialectica* 12 (1958), 280–7.

II. Significance of syntactic analysis

In accordance with para. 3 and following Herbrand's lead we do not formulate the results of syntactic analysis as consistency theorems, but as interpretations in the sense of [Kreisel 1958b].

13. Mathematical significance. To avoid repetition, we refer here to a recent discussion of this matter [Kreisel 1958b], in particular the application of an interpretation to the independence problem and other mathematical questions. To keep matters up to date we note that an interpretation in the sense of [Kreisel 1958b] of classical analysis in a quantifier-free *classical* theory of continuous functionals of finite type has been given whose constants are *recursive* continuous functionals [Kreisel 1959].

As a result we have a new

Open Problem. To give explicit characterizations (schemata) of the particular recursive functionals actually needed for the interpretation of one of the current systems of analysis.[16]

For independence proofs it is desirable to give alternative schemata.

Digression. A totally different type of syntactic study of classical analysis is suggested by so-called predicative enterprises, e.g. [Lorenzen 1955; Spector 1957; and Wang 1954]. It is evident that we do not get a model of classical analysis C simply by letting the variables of higher

nificance of this step specially for the Hilbert programme for arithmetic.

[15]I regard the use of infinite induction by Schütte and Lorenzen as an intermediate step corresponding to the use of intuitionistic arithmetic in Gödel's work: it is to be supplemented by an analysis by means of ordinals.

[16]An important contribution to this problem is made in Spector's paper [1962, section 10].

type[17] in a formula of C range over the sets of Wang's Σ_α for a fixed α[17]; for if α is not a limit number we conflict with the theorem of the least upper bound, and if α is a limit number we conflict with the existence of a non-denumerable set.[18] However, it seems promising to try this: given a proof in classical set theory to index the variables actually occurring in the proof by means of ordinals α so that the proof goes into a set of true statements in Spector's sense [1957]. This would not give an interpretation in the sense of [Kreisel 1958b] because the statement $(Ef)(g)A(f,g)$ is replaced by $(Ef_\alpha)(g_\alpha)A(f,g)$, with special α, from which $(Ef)(g)A(f,g)$ cannot in general be inferred, but nevertheless such an indexing would yield independence proofs if for certain formulae of classical set theory there is no indexing which yields true statements (if indices α less than an appropriate bound are used).

The situation seems not unlike Euclidean geometry. For, just as in analysis, the most natural conception of a point ignores the matter of naming the point, i.e. how the real number is represented or by what constructions the point is reached from given points. But if one wants to assert the impossibility of, say, ruler and compass constructions, one introduces coordinates of a suitably restricted kind. Actually, the analogy is not complete because the field of real square root extensions of the rationals is a model of the Euclidean geometry considered, i.e. without the general continuity axiom, whereas the class of Σ_α sets is not a model of classical analysis if α is recursive.[18]

14. Remark on the methods used in syntactic analysis. Most of the mathematical applications of syntactic analysis, whether achieved or projected, depend little on restrictions of the metamathematical methods, except perhaps for primitive recursive bounds as in [Kreisel 1958b: 165]. In [Łoś, Mostowski, and Rasiowa 1956] there is an attractive exposition of a non-finitist approach to the complex of Herbrand-type theorems[19] for predicate calculi. – It is perhaps fair to say that the most significant

[17] *Type* in the sense of the simple theory of types; α is called a *level*.

[18] More precisely: (*a*) For countable α, Σ_α does not satisfy the axioms of classical analysis A if, as in [Wang 1954], Σ_α contains an enumeration of the union of $\{\Sigma_\beta : \beta<\alpha\}$, possibly in addition to the usual ramified hierarchy (Σ_α^-). (*b*) For $\alpha<\omega_1$, i.e. recursive α, such an enumeration is explicitly definable in Σ_α^-. (*c*) Significant parts of A are realised in $\Sigma_{\alpha+\omega}^-$ (for all α) and in $\Sigma_{\omega_1}^-$, but e.g. the Cantor–Bendixson theorem is not realised in any Σ_β^-, $\beta\leqslant\omega_1$; for a survey [cf. Kreisel 1960b]. (*d*) There are denumerable α, α' ($\alpha'>\alpha$) such that, for $\beta>\alpha$, Σ_β^- satisfies A; for $\beta>\alpha'$, Σ_β^- satisfies classical set theory, and, in both cases, strong axioms of constructibility [Cohen 1963b]. Thus the ramified theory is useful for the syntactic study of some subsystems of A by (*c*), for the reduction of extensions of A *to* A by (*d*), but not for the study *of* A itself.

[19] It should be observed that, in contrast to mere consistency problems, Herbrand-type theorems remain significant even if no restriction is imposed on the metamathematical methods in contrast to the consistency "proof" of para. 8.

results in this area were first discovered through a good understanding of finitist or intuitionistic conceptions simply because certain results are evident from a constructive meaning of the formulae involved. But, at least once the results are known, one would expect to prove them more simply by full use of non-constructive methods. – The situation is different at present in the study of systems of arithmetic such as Z (cf. end of para. 8).[20] There is a temptation to regard these results as artificial because the systems studied are fragments of arithmetic (by incompleteness) and may be freaks to which freakish methods are well adapted. – I do not believe that this is a fruitful account of the position; on the contrary, I believe we are here faced with a very good problem of foundations. But I do not have a satisfactory answer: we have come back to the third problem of para. 4 (cf. para. 18).

15. Modified Hilbert programme. We take for granted Gödel's two incompleteness theorems and that finitist proof in its original sense does not go essentially beyond classical arithmetic [Hilbert 1932–5, 3: 212]; the latter would follow if the formulation of finitist proof sketched below is accepted. Thus it is necessary to use a hierarchy of constructive methods for the study (interpretation) of axiomatic systems. As we said above, we regard Herbrand's theorem as the paradigm for such an interpretation.[21]

The question is of course which aspects of Herbrand's theorem are essential for the present purpose, namely the understanding of transfinite symbols, and which are details, perhaps of importance elsewhere.

As in para. 11, according to Herbrand, $\mathfrak{A}$, say $(x)(Ey)(z)A(x,y,z)$, is provable in the predicate calculus, if and only if one of a sequence of quantifier-free A_n is provable, A_n being of the form

$$A\{\alpha, \varphi_1(f), f[\varphi_1(f)]\} \vee \ldots \vee A\{\alpha, \varphi_{p(n)}(f), f[\varphi_{p(n)}(f)]\}$$

Note in passing that quite trivially the sequence could be replaced by a single formula with a constructive metamathematical quantifier $(\exists n)$, namely $(\exists n)A(n)$, where $A(n)$ denotes A_n. (In the case of number theory and analysis we do not even need a metamathematical quantifier, but a

[20] Cf. [Kreisel 1958b] and footnote 7.

[21] In the sense of [Kreisel 1958b]. The mathematical use of interpretations for independence proofs, etc., is discussed elsewhere and is here taken for granted. Mostowski [1957b] and Tarski have protested against my use of the word "interpretation." There are two issues: (i) It clashes with Tarski's use of the word in [Tarski, Mostowski, and Robinson 1953]. (ii) Does it express at least one important meaning of the word in common use? – As to (i) there is a perfectly good word current for Tarski's meaning, namely model, and in case of doubt, one could use "formal" or "syntactic" model; also that book appeared much later than my first use of the term in [Kreisel 1951–2]. But (ii) raises a serious question if one is interested in the notion of a "reduction"; the discussion below was stimulated by Mostowski's criticism *lc*.

constructive existential quantifier $(\exists\varphi)$, φ ranging over a suitable class of recursive functionals.)

$\mathfrak{A}$ and A_n are in the close relation that $\mathfrak{A}$ can be proved from A_n by means of Herbrand's rules which have the property that it is *decidable*[22] for given $\mathfrak{A}$ and A_n, whether $\mathfrak{A}$ be so provable from A_n. What seems to me significant about this is that the logical relation between $\mathfrak{A}$ and A_n considered is essentially more elementary than the logical relations discussed, namely those of the (undecidable) predicate calculus. In this sense A_n expresses the full content of $\mathfrak{A}$.

Furthermore, as in Bernays' consistency theorem, if $\mathfrak{B}$ is provable from $\mathfrak{A}$ by means of transfinite symbols and $\mathfrak{A}$, interpreted as one of the A_n, is finitistically true, then so is some B_m.[23] – It is natural that a good understanding of the use of transfinite symbols should include the consideration of implications $\mathfrak{A} \vdash \mathfrak{B}$ since their purpose is not only to produce tautologies, but also, as we have said, to step from extra-logical axioms to theorems.

This analysis led very obviously to the notion of interpretation, in particular, in finitist (constructive) systems. The conditions given are intended to express what one would reasonably expect of an understanding of transfinite symbols by, or of a reduction to, finitist (constructive) means. Whether they do so, seems a proper subject for discussion by philosophers. I myself would apply the word *reduction* only to an interpretation of a system in one of its subsystems where the primitive notions and theorems of the latter are a subset of the former. – In particular, I do not call the set theoretical definitions of natural numbers or finiteness (cf. footnote 6) a reduction because, though we now have only the one primitive notion of a set, its content is not clearly comparable with that of the others. – Further I should require the reduction to be established by (sound) methods which can be formalized in the subsystem itself.

For me the "reduction" of primitive notions is not a matter of principle. A reduction does not eliminate them since merely to see that a proposed reduction is correct one has to start with the primitive notion considered anyway. But in practice such reductions can be extraordinarily fertile if for no other reason than that they permit the formulation of new questions: of $(x)(Ey)A(x,y)$ we ask only if it is valid or not, of $A[\alpha, t_1(\alpha)] \vee \ldots \vee A[\alpha, t_n(\alpha)]$ we ask what the complexity of the terms $t_i(\alpha)$ is. This is typical of the increased information contained in Herbrand's theorem. Another example is the arithmetization of the completeness theorem in [Hilbert and Bernays 1934–9: vol. 2] compared with the mere assertion of completeness. What kind of additional information is regarded as providing a satisfactory answer is a similar kind of question to the one about the significance of finitist proof; see also end of para. 18.

[22] Since in any application of Herbrand's rules the number of disjunction symbols or of free variables is reduced.

[23] This is really clear for an applied predicate calculus only, cf. in note on the pure predicate calculus in the remark below.

Open problem. For a consistently constructive interpretation of analysis (para. 13) it is necessary to prove constructively the existence of the continuous functionals of whatever schema is produced in the solution of the open problem of para. 13. This problem is analogous to the proof of existence of Gödel's functionals in para. 12.

16. Remark. It seems desirable to discuss the bearing of an *interpretation* on Brouwer's objection that consistency leaves open the possibility that some provable theorems are intuitively false. Remembering the difference between the intuitionistic and truth functional meaning of the logical constants one must use an appropriate translation of classical theorems before speaking of their being intuitively false. Gödel elucidated Brouwer's point most elegantly: in as much as consistency does not ensure ω-consistency, Brouwer was evidently right (even if one translates classical $(Ex)A(x)$ into intuitive $\sim(x)\sim A(x)$); on the other hand for classical number theory in particular Brouwer's objection did not arise because of the translation into Heyting's number theory. In the case of Herbrand's own interpretation the following points should be noted. Using Hilbert's terminology it is clear that quantifier-free formulae, i.e. those of the elementary calculus with free variables [Hilbert and Bernays 1934–9: vol. 2], are regarded as the "real" elements, others as ideal and in need of an interpretation. Now, Brouwer would certainly not accept this because on his interpretation of propositional formulae, the theorems of the classical propositional calculus are not valid; in particular, he interprets $A \vee B$ as: A is provable or B is provable; also he wishes to substitute for propositional letters incompletely defined propositions, e.g. propositions containing a parameter ranging over free choice sequences; under this interpretation $A \vee \sim A$ is evidently not valid. On the other hand, if a finitist had made Brouwer's objection, Gödel's translation into Heyting's arithmetic would not have been sufficient, and to answer it something like the extensions of Herbrand's theorem *Ic* would have been necessary. However, now the formulae of the classical propositional calculus could be regarded as "real" elements because for finitist propositions the truth-functional interpretation of the logical connectives is applicable.

We note in passing that just as there was a Hilbert programme for understanding classical transfinite machinery from a finitist point of view so there is an analogous programme for understanding intuitionistic machinery. Certainly as far as independence results are concerned, the latter is more rewarding if for no other reason than that it is less familiar.

Finitist proof

17. Work is in progress[24] on a characterization of finitist proofs in the usual sense: a formal system is described such that (i) each formal proof of the system is recognized as a finitist proof, and (ii) each formula in the notation of the system is asserted to be provable in the system if it is provable by finitist methods at all. The variables of the system are of two types (natural numbers and free function variables from the natural numbers to the natural numbers though the latter can be avoided), the constants are particular numerals, certain constant functionals and (finitist) proof predicates, the last two being introduced only after certain existential statements have already been established. The idea corresponds closely to what Hilbert imagined the whole mathematics to be like, namely an interplay between formal proofs and metamathematics [Hilbert 1932–5, 3: 174–5]. Finitist proofs constitute then the least class of proofs closed under the following condition (and containing a certain obvious minimum): (i) if a proof predicate has been shown by finitist methods to satisfy the relevant existential conditions then it is finitist too, and (ii) if $\mathrm{Prov}(n, m)$ is a finitist proof predicate already introduced, and if, with free variable n, $(\mathrm{E}p)\mathrm{Prov}[p, \ulcorner \mathrm{A}(0^{(n)})\urcorner]$ is established by a finitist proof, then so is $\mathrm{A}(n)$; – by Gödel's theorem this closure condition can be achieved only if no predicate of the system itself is both extensionally equivalent to the proof predicate of the whole system and also satisfies the conditions on a proof predicate imposed in Gödel's second undecidability theorem.

Evidently, there is no reason why the class of theorems should not be recursively enumerable: in fact, it is. A finitist could even conjecture that a particular enumeration gives precisely the class of finitistically provable theorems, e.g. in the notation of primitive recursive arithmetic, but this would be, for him, an empirical conjecture incapable of (finitist) proof.

Just as with the class of recursive functions, the only completeness properties of our class of proofs are certain closure properties, i.e. it is the least class with these closure properties.

The main open "problem" is to discover intuitively really convincing

[24] This has since been presented in the sketch [Kreisel 1960b], where the characterization of other informal notions of proof is also considered. – My interest in a definition of "finitist proofs" was reawakened by conversations with K. Gödel, 1955–7.

completeness properties, aided of course by more detailed information about the class itself, cf. para. 4 above.

As pointed out in para. 4, other classes of constructive proofs should be studied in addition to finitist proofs. In particular, as Bernays mentioned [1941: 151], the use of the first ϵ-number is intermediate between finitist and full intuitionist mathematics. It would be interesting to motivate the choice of some subclass of intuitionist proofs which includes the use, made e.g. in Ackermann [1940], of the first ϵ-number. (I have not done this to my own satisfaction even for the notion of finitist proof.)

18. General remarks. Since our understanding of the notion of "constructive proof" and of its special case "finitist proof" is not too detailed, the very meaning of Hilbert's programme is not too precise. However, one can summarize what parts of Hilbert's programme are settled on the basis of our partial understanding of these notions (cf. end of section 1). If one prefers a more formal approach, one would begin with a (partial) axiomatization of the notions of "constructive proof" or "finitist proof" which contains only trivial properties of these notions, and then investigate from what additional axioms for these notions the assertions below can be formally derived.

Hilbert's programme in the wide sense wanted to establish the "adequacy" of deductive formalisms for the representation of intuitive branches of mathematics. Their inadequacy (in the case of arithmetic) is established by Gödel's first incompleteness theorem on the basis of our partial understanding of the notions involved, namely the recursive enumerability of the set of theorems of any deductive formalism in the original sense; then there is an A such that $[(\ulcorner \mathrm{A} \urcorner \text{ is a formal theorem}) \rightarrow \mathrm{A}]$ is not formally derivable, and so the formalism cannot formally be proved to be sound if it is sound (for this A); further, it is inadequate, where by "adequacy" one means: if A then ($\ulcorner \mathrm{A} \urcorner$ is a theorem). It is interesting to observe that this inadequacy of the usual formalisms is connected with their surprising adequacy in another sense, namely the possibility of representing all recursive functions, at numerical arguments, by terms of the usual formal systems. As far as I know, before Gödel's work it was not even realized that all multiply recursive predicates could be defined in first order arithmetic, even when symbols for primitive recursive functions are added; although multiply recursive predicates were accepted as finitist. In fact, the schemata for recursions of higher type which Hilbert considered [1899b, 7th ed.: 295 and elsewhere] could, at least at first sight, be expected to "transcend" even analysis. On the other hand, as we know now, the proofs in present-day number theory can indeed be rather easily formalized in Hilbert's own formal system Z of arithmetic by means of the devices introduced by Gödel. So, while in Hilbert's days the empirical evidence (from "ordinary" number theory) for the adequacy of Z was slight, at the present time it would be overwhelming! It seems clear therefore that Hilbert's grounds for the feasibility of his programme must have rested on general philosophical considerations, perhaps the following: All that we "do" in mathematics (or: in thinking generally) is to operate with symbols, and that is all we "really" communicate to one another. (Hilbert sets great store by the "finiteness of our thoughts," both in principle [1932–5, 3: 187] and in proof theory, cf. para. 10 above.) As far as theorems of a transfinite character are concerned, we have proofs: their "truth" is metaphysical or poetic, and any reference to their truth must be "ultimately" reducible to assertions of formal derivability. The notion of mathematical truth can have no place in mathematics as we know it. – While it would be generally granted that this reflection itself is metaphysical or poetic it is convincing because in "ordinary" mathematics there is not the slightest hint of any practical use of the distinction between "truth" and "formal derivability" (even when rules of derivability are made explicit). The first thing to do, if one takes this distinction at all seriously, is to consider the assertion: ($\ulcorner A \urcorner$ is a formal theorem) $\rightarrow$ A. Expressing this assertion by means of an arithmetic formula is an essential step towards the incompleteness results. Also, the truth of each such formula is evident from the interpretation of the formal rules of proof and not from the combinatorial use of these rules. In other words, each such formula is obtained from the intended meaning of the formal systems considered, i.e. accepted on the basis of this meaning, without being formally implied by the given rules (unless A itself is so implied). Each such formula would be accepted as an axiom for arithmetic. Thus we have here an illustration of how one chooses axioms for formal systems from the intended meaning of the formal systems to be considered. The part of mathematical activity concerned with a good choice of axioms had no place in Hilbert's "official" conception of mathematics. If there is any real justification for calling Hilbert's approach "formalist" it is certainly this deficiency of Hilbert's official conception of mathematics and not his use of syntactic formulations in the foundations of mathematics (cf. para. 1).

Next consider Hilbert's programme in the narrow sense, namely to prove the consistency of the usual formalizations of mathematics by finitist means. Strictly speaking, he had a number of intermediate conjectures, too, such as the decidability of arithmetic or the somewhat peculiar assertion (*): it is consistent to assume that every problem of arithmetic is solvable. But a detailed discussion seems of little more than biographical interest, since these conjectures are either settled or genuinely ill formulated: if (*) means that one may consistenty add to formal arithmetic the statement:

> *For every closed formula* A *either* A *is a formal theorem or* $\sim$A *is a formal theorem,*

then the conjecture follows trivially from Gödel's result that one may even add consistently the assertion that the formal system is inconsistent. It is unlikely that Hilbert was satisfied by this manner of establishing his conjecture. Naturally, for the narrower programme an impossibility proof requires a more detailed analysis of the notions involved than for Hilbert's general programme. In particular, Gödel's second incompleteness theorem is not conclusive until reasons are adduced which show that all finitist theorems are included among the formal theorems of the system considered. As regards a characterization of finitist proof the sketch [Kreisel 1960b] leaves open the problem of upper bounds for the set of finitist theorems, and one may have to be content to obtain bounds in stages. It is, for instance, quite likely that more convincing arguments can be given, i.e. fewer assumptions on the notion of finitist proof are needed, for showing that all finitist functions are ordinal recursive of order $<\epsilon_0$ than, e.g. for showing that the consistency of first order arithmetic cannot be proved by finitist means. (In other words, the impossibility of a finitist proof for the ω-consistency of arithmetic may be genuinely easier to establish.) We may note in passing that current mathematics illustrates the use of different properties of intuitive intensional concepts in different theorems. Thus Gödel's first incompleteness theorem, which is concerned with Hilbert's general programme, uses substantially fewer properties of the intuitive notion of "representability" of (the syntactic) properties in a formal system than the second, which is concerned with the narrower programme. The first theorem requires that the set of theorems be representable, i.e. the formal system considered may be "identified" with its set of theorems (i.e. all we need know about it is its set of theorems), while the second requires certain (internal) properties of the proof relation to be formally derivable. This will be the case if one "identifies" the formal system with its set of production rules but not if one identifies it with the set of theorems since this set can be generated by different rules. (One could state this explicitly in the following formulation: *If a formula can be proved in* S *to express the consistency of* S, *and* S *is consistent, then this formula cannot be proved in* S; certain minimal conditions on the notion "expressing the consistency of S" are formulated in [Hilbert and Bernays 1934–9: vol. 2].)

It is to be emphasized that the remarks above do no more than assert a conviction that the notions of "constructive" and "finitist" are ripe for a systematic study. The reformulations of known theorems given above indicate that our conception of these notions is coherent because these reformulations are not forced.

We conclude with some pragmatic remarks on the choice of a significant class of constructive methods which goes beyond finitist mathematics. Here one has to recognize the following difference between the untutored notion of constructivity of the "ordinary" mathematician (interested in constructivity) and the one that seems to be forced on one if one tries to be coherent. Naïvely, "constructive" is applied to procedures or instructions for manipulating finite configurations (as in the case of finitist mathematics), but not to operations on objects of higher type or logical operations (e.g. on proofs with meaning, as in the case of intuitionistic mathematics). The naïve conception does not get into difficulties in familiar parts of mathematics, because, once constructive definitions are given, the proofs are in general quite unproblematic, cf. footnote 5. On the basis of this experience the mathematician is ill-prepared to judge the constructivity of, e.g. definitions by means of recursion of the form $f(n)=g\{n,f[\tau(n)]\}$ if $\tau(n)\prec n$ and $f(n)=g(n,0)$ if $\tau(n)\not\prec n$, where g,τ are constructive functions, $\prec$ a well-founded constructive ordering (but may not have been proved to be well founded by "constructive means"). This is precisely the situation in several consistency proofs for arithmetic [e.g., Ackermann 1940]. If one thinks in terms of (idealized) machines, the above definition is a proper instruction, provided $g,\tau,\prec$ are constructive, but the *proof* of well-ordering is irrelevant. It is likely that, on this naïve conception, recursion on (the usual orderings whose ordinal is) the first ϵ-number would be more evidently "constructive" than Gödel's simple functionals quoted in para. 12 above, which involve variables of higher type.

On the other hand, both philosophically and pragmatically, there is strong evidence that a more coherent theory of constructivity is possible if this notion is applied also to logic (inference), in particular to the interpretation of logical constants. For instance, philosophically, it is clear that there is a gap in accepting as constructive the definitions by transfinite recursion of the preceding paragraph if no condition is imposed on the proof of well-ordering: one certainly could not claim that the definition has been justified on constructive principles. Pragmatically, i.e. for the purpose of obtaining, in a systematic manner, well-rounded mathematical theories, the use of the primitive notion of a constructive proof has definitely been fruitful. One example is Gödel's use of intuitionistic logic as an intermediary in the consistency proof of arithmetic (cf. para. 12), another is the use of generalized inductive definitions based on intuitionistic logic, which explains the remarkable fact that many functions

defined on inductive sets (e.g. Church–Kleene ordinal notations) are recursive: this is mysterious from the non-constructive approach to recursion theory because the inductive sets are highly non-recursive, and one sees no reason why definitions by transfinite induction on such sets should lead to recursive objects. But, also the following more isolated example is instructive.

Suppose we ask for a modification of classical arithmetic which leaves the constructive theorems (of the form $(x)(Ey)A(x,y)$ with quantifier-free A) unchanged, but makes every prenex theorem recursively satisfiable. This would express more or less what the naïve constructive mathematician described above might be after. In such a modification the class of quantifier-free theorems is the same as in classical arithmetic, while the class of theorems containing quantifiers is changed. Who would have thought of modifying only the rules for the propositional connectives, and not for quantifiers, which is precisely what is done in intuitionistic mathematics. However, if one has the concept of constructive proof and Heyting's interpretation of the logical concepts, this step is inevitable. Moreover, while the general problem (of finding rules of proof valid for the constructive interpretation of the logical constants) has a high degree of determinateness,[25] the naïve constructivist's question certainly does not have a unique answer. Thus $[\sim p \to (q \vee r)] \to [(\sim p \to q) \vee (\sim p \to r)]$ could be added to the propositional axiom schemata of Heyting's arithmetic, or, again, for primitive recursive $A(x)$, $\sim (x)A(x) \to (\exists x)\sim A(x)$; both these additions are nonconservative, since the former schema is not a theorem of Heyting's propositional calculus and the latter is deducible in his arithmetic only if $(\exists x)\sim A(x) \vee (x)A(x)$ is too.

More generally, even if one is primarily interested in developing systematically a smooth-running formalism, it is necessary to develop a coherent philosophical notion of constructivity. The conception itself (which applies to functions, proofs, etc.) has a degree of generality which transcends any mathematical treatment, but it is the business of pure mathematics to develop it within suitable specific contexts. In this respect it is like the notions of structure, truth, or proof. It may be remarked that I myself have come to recognize the need for, and usefulness of, a (non-finitist) notion of constructive proof only very reluctantly.

Completeness of the predicate calculus

19. I wish to emphasize two points of interest of the completeness problem; namely it illustrates (i) the mathematical treatment of a concept of

[25]Even though the possibility of a (provably) complete formalization is doubtful.

great generality in a specific context, as mentioned in the preceding paragraph, (ii) the limitations of a purely finitist metamathematics.

If $A(P_1, \ldots, P_n)$ is a (closed) formula of the predicate calculus whose predicate symbols are $P_1, \ldots, P_n$, then it is valid if and only if for every domain D of individuals and predicates P_i^o, $1 \leqslant i \leqslant n$, defined in D, $A(P_i^o, \ldots, P_n^o)$ is true on the usual interpretation of the logical symbols. The calculus is complete if every valid formula is provable in it. The reference to arbitrary domains here is a typical case of philosophical generality, reminiscent of the paradoxes. It would be disturbing if the paradoxes led to the formalization of logic and the justification of the formalization led back to the very notions involved in the paradoxes! Now, if one merely wishes to show that every theorem of the calculus is valid, one considers any particular set D, presumed to be well defined, and follows out the rules of the calculus with quantifiers ranging over this set; there is no generation of new sets, which is the typical step in the paradoxes. But, if one shows that every valid formula is provable, the premise itself involves quantification over all sets. As long as one has a precisely limited notation, one may expect to sharpen the result by restricting the sets for which the formula has to be valid. Gödel shows that for each $A(P_1, \ldots, P_n)$ there is a *single* set of predicates P_i^A defined over the domain D_o of natural numbers such that instead of the validity of A we merely need the truth of $A(P_1^A, \ldots, P_n^A)$ in D_o.

From a finitist point of view even this is not enough since in general the P_i^A cannot be chosen to be constructive at all. There are two finitist versions of the result, namely (a) if $A(P_1^A, \ldots, P_n^A)$ is provable in Z then $A(P_1, \ldots, P_n)$ is provable in the predicate calculus, or the stronger form (b) $A(P_1^A, \ldots, P_n^A) \rightarrow$ "$A(P_1, \ldots, P_n)$ is provable in the predicate calculus," the implication being provable in Z for each A where the statement in inverted commas denotes some natural arithmetization of itself. – It should be remembered that these versions are not what Hilbert originally required [1899b, 7th ed.: 322] namely that if $A(P_1, \ldots, P_n)$ is not provable in the predicate calculus, $\neg A(P_1^A, \ldots, P_n^A)$ should actually be provable in Z. There is no recursively enumerable extension of the predicate calculus with this property. (Hilbert evidently assumed Z to be complete and therefore did not distinguish between these versions which are equivalent on his assumption.)

Of course, on occasions the extra information contained in the finitist versions is needed. But the finitist point of view is simply not appropriate for discussing the general problem of completeness.

Postscript (Autumn 1978) to "Hilbert's Programme"

After twenty years it is, fortunately, necessary to add bibliographical references to bring the technical part of the article up to date. This is done in short notes that refer to the relevant chapters of the article (without giving the verse). In addition I wish to stress two general changes in my views over the last twenty, and especially the last ten, years.

First of all, the article seems far too conciliatory on the *finitist philosophy of mathematics*. Obviously this philosophy gets its sting from its negative side, the rejection of non-finitist methods and particularly non-finitist formulations, and much less from its positive side, the mere use of finitist methods. This is so because, quite often, a finitist proof tells us something we want to know that its competition, for example, a slicker non-finitist proof, does not. Thus no grand "philosophy" is behind the almost universal preference for a proof of convergence of a^n $(0 \leqslant a \leqslant 1)$ that computes bounds for the rate of convergence over its competitor; the latter notes monotonicity and boundedness of the sequence a^n, and then appeals to the principle of the least upper bound (cf. also Hilbert's example in the last paragraph but one of §4). As I see things now, it is little short of an evasion to call the switch in §3 from Hilbert's *consistency problem* (which concerns only proofs of purely universal statements) to an *interpretation* (of logically compound formulae) a "critique of detail": this logically fruitful and quite decisive switch draws attention away from the consequences of rejecting non-finitist proofs, for example, the perfectly sensible consistency proof mentioned at the beginning of §8. Behind such proofs there is very real progress, provided, of course, that there is a clear description of the structures for which the axioms considered are true. (For example, the clear description of segments of the cumulative hierarchy of sets has done more for the removal of genuine doubts than all finitist consistency proofs put together.) Presumably, mathematical practice has not been affected much by finitist doctrine: For one thing, there are plenty of interesting problems within those parts of mathematics that are simply *intended* to be finitist. But foundational research has unquestionably been affected negatively, particularly – and ironically – proof theory itself. ("Ironically" because this branch of mathematical logic was created by Hilbert for the sake of a finitist philosophy.) The mechanism at work is quite simple. Consciously or unconsciously, the bulk of publications in proof theory stress those results that, like the consistency of currently used principles, are easy or trivial unless one adopts (the negative side of) the finitist doctrine. For this very reason members of the silent majority, not only of mathematicians but of logicians, are put off by (the dullness of these results in) proof theory.

As is usual in such circumstances, the doctrinaire finitists have come to feel isolated and to think of the silent majority as lacking that Higher Sensibility that is needed to appreciate finitist doubts, though, as somebody said, those doubts are more dubious than what is being doubted.

The second major defect of the article is, so to speak, accidental, and related to the following circumstance. During the years 1956–8, I managed to do something of interest both with the notion of finitist proof itself (§17), by use of so-called autonomous progessions, and the more general notion of constructive proof (§18), for example, in connection with the completeness of Heyting's systems. As I see matters now, I was overimpressed by the fact that *anything* precise could be done with these notions, and even more by the fact that the work was most *satisfaisant pour l'esprit*. (The euphoria lasted for some time because closely related considerations worked for the notion of predicative proof and led to refinements of then-current notions concerning ω-models and generalizations of recursion theory, and finally to the proper choice of languages with infinitely long formulae). What I failed to do was to consider in detail any genuine alternative, so to speak competing, categories (of proofs) that might serve the same general purpose as the traditional categories (finitist, constructive, predicative). In particular, I failed to pursue the issue, stated emphatically in §18, whether one should extend to proofs, that is, to logical inferences, the more naïve categories of operations and definitions. In this connection recursion theory had already provided a convincing analysis of (hereditarily) *finite* operations. Sure, from a finitist point of view it is obviously coherent to restrict also the proofs that establish that a proposed rule for an operation is well-defined. But as I see things now – and contrary to §18 – there was little evidence that this finitist aim would be rewarding (compare it to the aims of alchemy, which are only occasionally rewarding for chemistry); all the more since, by §4, the restriction in question to finitist proofs was not expected to increase reliability.

Looking back, I remember a number of such alternative categories that had struck me before 1958 and that have since turned out to lend themselves to logical study. (It is of little interest – even to me personally! – whether I "should" have studied them rather than finitist or intuitionist principles of proof.) For example, there was the matter of *structural complexity* of proofs. This cuts across the traditional categories. It is critical for reliability, since if there is no doubt about the validity of principles of proof, the decisive factor is the probability of error in applying those (correct) principles: Here the complexity of individual proofs is relevant. Evidently, there is not only one (simple) measure of complexity,

no more than in the case of physical objects: Their complexity, in the sense of difficulty of grasping them, involves length, volume, weight, shape, even chemical composition, electric charge, and so on. The most one can hope is that a *few* measures are significant for, that is determine, *many* issues that arise in the course of nature. As it happens, there are a couple of abstracts on *relative consistency proofs* (Kreisel 1958c: 109–10 and Kreisel 1976a: 285–6) that illustrate the development strikingly. The first, stressing principles of *proof* only, provides a mere counterexample; the second, stressing *operations* (relating hypothetical proofs of inconsistency in the two systems), gives a satifactory solution.

As another example that certainly fits in with finitist preoccupations and with the thoughts on incompleteness at the end of §14: By and large I neglected subdivisions *within* finitist mathematics or first order arithmetic, and barely considered the mathematical significance of such subdivisions in the sense of §13. Very soon after those preoccupations ended I noticed that van der Waerden's theorem on arithmetic progressions (which I have known very well since my student days) presented an excellent candidate for using *primitive* recursion: All the known proofs of the theorem use double recursion (footnote 7) to bound N in terms of k and l:

> If $\{1,\ldots,N\}$ are divided into k classes, at least one class contains an arithmetic progession of length l.

In contrast, derivability of van der Waerden's theorem in first order arithmetic is obvious and of no interest. In short, we have a new criterion for the choice of formal systems which are rewarding to study metamathematically, simply by concentrating on *particular* theorems. This observation was, as a matter of empirical fact, overlooked in §14 and the first paragraph of §15, where the superiority of more traditional logical criteria is tacitly assumed.

In fairness it should be added that, even with present experience, we can find very few corners in the mathematics of the 1950s that lend themselves to the type of logical analysis envisaged (and set out in Kreisel 1958b), particularly, the unwinding of proofs of Π_2^0 theorems in terms of the rate of growth of bounds, let alone of Π_1^0 theorems. This has changed significantly only in the seventies.

The general point of view of this Postscript is presented in a very condensed style in "What Have We Learned from Hilbert's Second Problem?" (Kreisel 1976b). [Added in autumn of 1982:] A leisurely presentation of the general point of view, in a broad context, is to be found in my obituary of Kurt Gödel (Kreisel 1980).

Notes

§4. Not only here, in the third paragraph, but throughout the article I make far too heavy weather of "derivability conditions" on the predicates and functions that are used to code arithmetically syntactic notions and operations concerning *given* formal rules $\mathfrak{F}$. In particular, I associated these conditions, in the sense of Hilbert and Bernays (1934–9, 2: 286) with certain differences between Gödel's and Henkin's self-referential sentences (I am not, resp. I am provable) which I had noticed in "On a Problem of Henkin's" (Kreisel 1953b: 405–6). But starting with (Kreisel 1962b: 243–6) I stressed the existence of *canonical representations* of syntactic notions, provided of course one has made up one's mind on the data used to determine $\mathfrak{F}$ (for example, rules in the style of Post), a perfect parallel to, say, the familiar algebraic representation of geometric notions. In particular, the representation is demonstrably unique up to isomorphism (or: equivalence in the case of predicates) once the notions to be represented are sufficiently analyzed axiomatically. Naturally, certain "brutal" questions may be quite insensitive to the choice of data; for example, recursive decidability for which the "abstract" set of theorems of $\mathfrak{F}$ is adequate – and equally evidently this is not so for proofs of the consistency of $\mathfrak{F}$. The point to remember is that, for example, in Gödel's second theorem, it is not vagaries of representations of a given $\mathfrak{F}$ that are at issue, but so to speak vagaries among formal rules. Specifically, unorthodox "representations" of the kind used by Rosser to improve Gödel's first theorem, for $\mathfrak{F}$, are canonical for different rules $\mathfrak{F}_R$ (where $\mathfrak{F}$ and $\mathfrak{F}_R$ have the same set of theorems if $\mathfrak{F}$ happens to be consistent). And one of the "derivability" conditions (no. (iii), demonstrable completeness for Σ_1^0 sentences) is *not* satisfied by that *canonical* representation. The matter is put straight by use of an example in a leisurely discussion in the second edition of Hilbert and Bernays (1968–70, 2: 298–301) and discussed fully in my joint paper with G. Takeuti, "Formally Self-referential Propositions for Cut-Free Classical Analysis and Related Systems" (Kreisel and Takeuti 1974: App. 3).

§5. The use of the completeness theorem (for predicate logic) to establish incompleteness, in the last sentence, refers to "Note on Arithmetic Models..." (Kreisel 1950: 265–85). In the meantime, I noticed how to extend the argument to get a "model-theoretic" proof of Gödel's second incompleteness theorem, naturally, for classical systems (cf. end of footnote 43 on p. 383 of Kreisel 1968; for a detailed exposition, see Smorynski 1977).

§6, *especially footnote 5*. In the last few years, non-constructive proofs have been given of (constructive) Π_2^0 theorems. But it can hardly be claimed that they were helped by "closer study of... constructive aspects of non-constructive methods." Logicians are particularly attracted by a variant RT_A, of Ramsey's theorem RT on partitions, which is *known* not to be provable in first order arithmetic (and thus not finitistically provable in the sense of Kreisel 1960a; cf. Paris and Harrington 1977 on "mathematical incompleteness"). But the variant was discovered through a combination of experience in model theory and the partition calculus for "large" cardinals. Actually, less than ten years ago, I was led – by the traditional preoccupation with provability, rather than with particular proofs – to misinterpret a fact that had very much struck me at the time; specifically, as shown by Jockusch (1972: 268–80), the infinite version RT_∞ of RT is not arithmetic and, certainly, by far the easiest proof of (any variant of) RT which combines RT_∞, and a compactness argument is not arithmetic either. It just so happens that RT also has a proof in a fragment of primitive recursive arithmetic. But because of the latter, I dropped closer study of the easy proof of RT via RT_∞. In any case, in the last couple of years, non-constructive proofs of Π_2^0 theorems have been discovered that use principles that go far beyond first order arithmetic, specifically, proofs using so-called generalized inductive definitions familiar from the theorem of Cantor–Bendixson in descriptive set theory. Also – and in contrast to RT_A – these new proofs are very much in the "mainstream" of mathematics (see, for example, Furstenberg 1977: 204–56). [Added Autumn 1982:] More on the unwinding of these proofs is to be found in Kreisel (1982: 50).

§8. A more detailed description of the "new type of analysis" can be found in my review (Kreisel 1962c: 250–5) of *Beweistheorie* (Schütte 1960a). But while it is true that, in principle, this new analysis does not restrict the means of independence proofs, in practice so far it has not enlarged them either! In particular, we now have perfectly good model theoretic proofs of the independence results first discovered by the new (proof theoretic) analysis. *Warning:* Footnote 7 is very much restricted to *constructive* mathematics! Specifically, there are very manageable non-recursive (non-standard) models defined by suitable ultraproducts.

§9. The blithe reference to "a practical decision method" is almost empty, at least when applied to the – full first order – theories that were usually considered in the fifties. None of those theories has a practical decision method, and the problem is to select classes of formulae that do:

By now, the standard example comes from the theory of diophantine equations

$$(\exists x_1 \epsilon \omega) \ldots (\exists x_n \epsilon \omega)[p(x_1, \ldots, x_n) = 0]$$

where the polynomials p are selected according to geometric properties of varieties: $p=0$, defined in finite fields! (Cf. also Meyer 1975: 132–54.)

§*10, especially footnote 10.* Hilbert's ϵ-calculus has been neglected during the last twenty years except for a few asides. At one extreme, it is not known (even without restriction on the metamathematical methods) whether the particular substitution method for analysis proposed in Hilbert and Bernays (1934–9, vol. 2) converges (cf. Kreisel 1965: 168, 3.351). At another extreme, for the set theoretic formalism with primitives: membership, union, pair, empty set, there is a primitive recursive method for deciding whether any formula in the ϵ-formalism can be realized (by means of hereditarily finite sets; cf. Ville 1971: 513–16, extended by Gogol 1978: 289–90). And though no doubt any finite set of "critical" ϵ-formulae and axioms of current set theory can be realized, this fact cannot be proved in set theory.

§*11.* Knowledge of cut-free systems has improved immensely during the last twenty years. (a) The model-theoretical proof of cut elimination, mentioned in the second paragraph (apparently for the first time), has been refined so as to be formalized in a fragment of arithmetic. Constructive analysis of this fragment yields an α-recursive rule ($\alpha < \omega^{\omega^\omega}$). (b) The notion of *semi-valuation* introduced for type theory by K. Schütte (1960: 305–26) extends naturally to predicate logic. Though, by (a), there is no (extensional) difference between validity in *all* total, resp. semi-valuations, there is a difference if valuations of suitably *restricted complexity* $\mathfrak{C}$ are considered; equivalently, if the category of proof trees without any infinite paths is extended to those without any path $\epsilon\mathfrak{C}$ (Kreisel, Mints, and Simpson 1975: 38–131). (c) Ibid. The 'cogent motivation' mentioned in para. 5 of §11 is made explicit: the (model-theoretic) motive is to construct a tree T_A of formulae that codes *all* countable valuations in which A is false; and the rules (of inference) reverse the rules needed for constructing T_A.

§*11, footnote 13.* Herbrand's own publication has been found defective (cf. B. Dreben, P. Andrews, S. Aanderaa 1963: 699–706).

§*12.* The use of infinite induction, that is, of the ω-rule has considerably more significance than is suggested by footnote 15 (cf. Kreisel 1976c: 177–223).

Concerning footnote 14: the (quite essential) use of function variables presents an essential departure from Herbrand's *intentions* whose conception of finitist proof did not allow such variables, and led to his complicated reformulation.

§*13.* Despite the heading, there is an occasional bias toward logical rather than mathematical applications! Specifically, an important relation between the rate of growth of bounds and formal derivability of Π^0_2 formulae from true Π^0_1 sentences is emphasized (Kreisel 1958b: 159, ii). But – in contrast to the Postscript above – I thought of using this relation for establishing the logical property of underivability.

§*14.* The doubts about the use of proof theoretic methods for analyzing the set of theorems prov*able* in fragments of arithmetic were more than justified, as explained in the Postscript and the note concerning §8. The methods have permanent value for analyzing, for example, unwinding *proofs*.

§*15.* The doubts (still unresolved, incidentally) about an adequate (general) analysis of such aims as "understanding by or reduction to limited means" and about a proper choice of limitation do not affect the *alternative* to traditional foundations mentioned at the end of the Postscript, where our starting point is a proof of a particular theorem and we know what more we want to know about it. The *value* of a foundational analysis seems to me to be properly measured by the frequency of particular cases where we don't know this by the light of nature.

§*17, footnote 24.* The exposition of the ideas of §17 given in Kreisel 1965: 171–2 is better than the sketch in Kreisel 1960a.

§*18.* The quite fundamental, and generally neglected, distinction made here, between Hilbert's programme in the wide sense (adequacy of formal systems for representing reasoning) and in the narrow sense (consistency of particular systems), is stated more forcefully at the beginning of "A Survey of Proof Theory" (Kreisel 1968: 321–88). In particular, Gödel's *first* incompleteness theorem is sufficient to refute the programme in the wide sense, especially, if one adds Hilbert's requirement of a "final solution" of foundational problems by mathematical means, mentioned in the last paragraph of §6 of the text. The comments on "representability" of syntactic properties at the end of paragraph 3 of §11, on Hilbert's programme in the narrow sense, are superseded by the note above to §4. In particular, using the type of formal rules called $\mathfrak{F}_R$ (see

p. 234) we get a system, which has (i) exactly the same *derivations*, not only the same theorems as, say, first order arithmetic, say, Z,

> can prove (ii) its own consistency, but for example,
> cannot prove (iii) its completeness for Σ_1^0 sentences

(whereas (iii) demonstrably holds for Z). The difference between $\mathfrak{F}_R$ and Z is that the procedure for checking derivations is different, involving so to speak a general comparison with background knowledge, a look at theorems proved "earlier."

Concerning footnote 25, there is a quite extensive literature beginning with a convincing use of topological "interpretations" to establish deductive completeness of Heyting's system of propositional logic, specifically, for predicates of what are nowadays called lawless sequences (Kreisel 1958d: 369–88). The most recent result in this direction covers the fragment $\{\wedge, \vee, \rightarrow, \forall, \exists\}$ (without negation); cf. Theorem 13 in *Elements of Intuitionism* (Dummett 1977, p. 288). On the negative side, the assumption that all constructive number theoretic functions are recursive (that is, Church's thesis extended to intuitionistic, not only mechanical rules) implies that the set of constructively valid formulae in $\{\neg, \wedge, \vee, \rightarrow, \forall, \exists\}$ is not recursively enumerable (for exposition, cf. A. S. Troelstra 1977b: 39–58). For a connected account of the variants of Heyting's rules mentioned in the last paragraph but one of §18, see the axiomatizations of sentences (demonstrably) valid for various realizability and functional interpretations in the monograph of A. S. Troelstra 1973.

§19. The version (a), in the last paragraph but one, is nowadays called "maximality" of predicate logic for (schemata in) Z, since the paper by Dana S. Scott, "Extending the Topological Interpretation to Intuitionistic Analysis II" (1970). An example showing that (b) is 'stronger' is provided by Heyting's predicate calculus, which is maximal for Heyting's arithmetic but, by the note concerning §18, not complete. Put simply, we have maximality because we have forgotten the same valid (logical) rules in pure predicate calculus and in the calculus applied to arithmetic.

A general account of completeness, in terms of so-called informal rigor, which is considered to be readable, is in Kreisel 1967: 138–71.

PART II

The existence of mathematical objects

Empiricism, semantics, and ontology[1]

RUDOLF CARNAP

1. The problem of abstract entities

Empiricists are in general rather suspicious with respect to any kind of abstract entities like properties, classes, relations, numbers, propositions, etc. They usually feel much more in sympathy with nominalists than with realists (in the medieval sense). As far as possible they try to avoid any reference to abstract entities and to restrict themselves to what is sometimes called a nominalistic language, i.e., one not containing such references. However, within certain scientific contexts it seems hardly possible to avoid them. In the case of mathematics, some empiricists try to find a way out by treating the whole of mathematics as a mere calculus, a formal system for which no interpretation is given or can be given. Accordingly, the mathematician is said to speak not about numbers, functions, and infinite classes, but merely about meaningless symbols and formulas manipulated according to given formal rules. In physics it is more difficult to shun the suspected entities, because the language of physics serves for the communication of reports and predictions and hence cannot be taken as a mere calculus. A physicist who is suspicious of abstract entities may perhaps try to declare a certain part of the language of physics as uninterpreted and uninterpretable, that part which refers to real numbers as space-time coordinates or as values of physical magnitudes, to functions, limits, etc. More probably he will just speak about all these things like anybody else but with an uneasy conscience, like a man who in his everyday life does with qualms many things which are not in accord with the high moral principles he professes on Sundays. Recently the problem of abstract entities has arisen again in connection with semantics, the theory of meaning and truth. Some semanticists say that certain expressions designate certain entities, and

Reprinted with the kind permission of the author and publishers from Rudolf Carnap, *Meaning and Necessity*, 2nd ed. (Chicago: The University of Chicago Press, 1956), pp. 205–221, and from *Revue Internationale de Philosophie*, vol. 4 (1950), pp. 20–40. The slightly modified version that was printed in *Meaning and Necessity* appears here.

[1] I have made here some minor changes in the formulations to the effect that the term "framework" is now used only for the system of linguistic expressions, and not for the system of the entities in question.

among these designated entities they include not only concrete material things but also abstract entities, e.g., properties as designated by predicates and propositions as designated by sentences.[2] Others object strongly to this procedure as violating the basic principles of empiricism and leading back to a metaphysical ontology of the Platonic kind.

It is the purpose of this article to clarify this controversial issue. The nature and implications of the acceptance of a language referring to abstract entities will first be discussed in general; it will be shown that using such a language does not imply embracing a Platonic ontology but is perfectly compatible with empiricism and strictly scientific thinking. Then the special question of the role of abstract entities in semantics will be discussed. It is hoped that the clarification of the issue will be useful to those who would like to accept abstract entities in their work in mathematics, physics, semantics, or any other field; it may help them to overcome nominalistic scruples.

2. Linguistic frameworks

Are there properties, classes, numbers, propositions? In order to understand more clearly the nature of these and related problems, it is above all necessary to recognize a fundamental distinction between two kinds of questions concerning the existence or reality of entities. If someone wishes to speak in his language about a new kind of entities, he has to introduce a system of new ways of speaking, subject to new rules; we shall call this procedure the construction of a linguistic *framework* for the new entities in question. And now we must distinguish two kinds of questions of existence: first, questions of the existence of certain entities of the new kind *within the framework;* we call them *internal questions;* and second, questions concerning the existence or reality *of the system of entities as a whole,* called *external questions.* Internal questions and possible answers to them are formulated with the help of the new forms of expressions. The answers may be found either by purely logical methods or by empirical methods, depending upon whether the framework is a logical or a factual one. An external question is of a problematic character which is in need of closer examination.

The world of things. Let us consider as an example the simplest kind of entities dealt with in the everyday language: the spatio-temporally ordered system of observable things and events. Once we have accepted the thing language with its framework for things, we can raise and answer internal questions; e.g., "Is there a white piece of paper on my

[2]The terms "sentence" and "statement" are here used synonymously for declarative (indicative, propositional) sentences.

desk?", "Did King Arthur actually live?", "Are unicorns and centaurs real or merely imaginary?", and the like. These questions are to be answered by empirical investigations. Results of observations are evaluated according to certain rules as confirming or disconfirming evidence for possible answers. (This evaluation is usually carried out, of course, as a matter of habit rather than as a deliberate, rational procedure. But it is possible, in a rational reconstruction, to lay down explicit rules for the evaluation. This is one of the main tasks of a pure, as distinguished from a psychological, epistemology.) The concept of reality occurring in these internal questions is an empirical, scientific, non-metaphysical concept. To recognize something as a real thing or event means to succeed in incorporating it into the system of things at a particular space-time position so that it fits together with the other things recognized as real, according to the rules of the framework.

From these questions we must distinguish the external question of the reality of the thing world itself. In contrast to the former questions, this question is raised neither by the man in the street nor by scientists, but only by philosophers. Realists give an affirmative answer, subjective idealists a negative one, and the controversy goes on for centuries without ever being solved. And it cannot be solved because it is framed in a wrong way. To be real in the scientific sense means to be an element of the system; hence this concept cannot be meaningfully applied to the system itself. Those who raise the question of the reality of the thing world itself have perhaps in mind not a theoretical question as their formulation seems to suggest, but rather a practical question, a matter of a practical decision concerning the structure of our language. We have to make the choice whether or not to accept and use the forms of expression in the framework in question.

In the case of this particular example, there is usually no deliberate choice because we all have accepted the thing language early in our lives as a matter of course. Nevertheless, we may regard it as a matter of decision in this sense: we are free to choose to continue using the thing language or not; in the latter case we could restrict ourselves to a language of sense-data and other "phenomenal" entities, or construct an alternative to the customary thing language with another structure, or, finally, we could refrain from speaking. If someone decides to accept the thing language, there is no objection against saying that he has accepted the world of things. But this must not be interpreted as if it meant his acceptance of a *belief* in the reality of the thing world; there is no such belief or assertion or assumption, because it is not a theoretical question. To accept the thing world means nothing more than to accept a certain form of language, in other words, to accept rules for forming statements and for testing, accepting, or rejecting them. The acceptance of the thing language leads, on the basis of observations made, also to the acceptance, belief, and assertion of certain statements. But the thesis of the reality of the thing world cannot be among these statements, because it cannot be formulated in the thing language or, it seems, in any other theoretical language.

The decision of accepting the thing language, although itself not of a cognitive nature, will nevertheless usually be influenced by theoretical knowledge, just like any other deliberate decision concerning the acceptance of linguistic or other rules. The purposes for which the language is intended to be used, for instance, the purpose of communicating factual knowledge, will determine which factors are relevant for the decision. The efficiency, fruitfulness, and simplicity of the use of the thing language may be among the decisive factors. And the questions concerning these qualities are indeed of a theoretical nature. But these questions cannot be identified with the question of realism. They are not yes-no questions but questions of degree. The thing language in the customary form works indeed with a high degree of efficiency for most purposes of everyday life. This is a matter of fact, based upon the content of our experiences. However, it would be wrong to describe this situation by saying: "The fact of the efficiency of the thing language is confirming evidence for the reality of the thing world"; we should rather say instead: "This fact makes it advisable to accept the thing language".

The system of numbers. As an example of a system which is of a logical rather than a factual nature let us take the system of natural numbers. The framework for this system is constructed by introducing into the language new expressions with suitable rules: (1) numerals like "five" and sentence forms like "there are five books on the table"; (2) the general term "number" for the new entities, and sentence forms like "five is a number"; (3) expressions for properties of numbers (e.g., "odd", "prime"), relations (e.g., "greater than"), and functions (e.g., "plus"), and sentence forms like "two plus three is five"; (4) numerical variables ("*m*", "*n*", etc.) and quantifiers for universal sentences ("for every *n*,...") and existential sentences ("there is an *n* such that...") with the customary deductive rules.

Here again there are internal questions, e.g., "Is there a prime number greater than a hundred?" Here, however, the answers are found, not by empirical investigation based on observations, but by logical analysis based on the rules for the new expressions. Therefore the answers are here analytic, i.e., logically true.

What is now the nature of the philosophical question concerning the existence or reality of numbers? To begin with, there is the internal question which, together with the affirmative answer, can be formulated in the new terms, say by "There are numbers" or, more explicitly, "There is an *n* such that *n* is a number". This statement follows from the analytic statement "five is a number" and is therefore itself analytic. Moreover, it is rather trivial (in contradistinction to a statement like "There is a prime number greater than a million", which is likewise analytic but far from trivial), because it does not say more than that the new system is not empty; but this is immediately seen from the rule which states that words like "five" are substitutable for the new variables. Therefore nobody who meant the question "Are there numbers?" in the internal sense would either assert or even seriously consider a negative answer. This makes it plausible to assume that those philosophers who treat the question of the existence of numbers as a serious philosophical problem and offer lengthy arguments on either side do not have in mind the internal question. And, indeed, if we were to ask them: "Do you mean the question as to whether the framework of numbers, *if* we were to accept it, would be found to be empty or not?", they would probably reply: "Not at all; we mean a question *prior* to the acceptance of the new framework". They might try to explain what they mean by saying that it is a question of the ontological status of numbers; the question whether or not numbers have a certain metaphysical characteristic called reality (but a kind of ideal reality, different from the material reality of the thing world) or subsistence or status of "independent entities". Unfortunately, these philosophers have so far not given a formulation of their question in terms of the common scientific language. Therefore our judgment must be that they have not succeeded in giving to the external question and to the possible answers any cognitive content. Unless and until they supply a clear cognitive interpretation, we are justified in our suspicion that their question is a pseudo-question, that is, one disguised in the form of a theoretical question while in fact it is non-theoretical; in the present case it is the practical problem whether or not to incorporate into the language the new linguistic forms which constitute the framework of numbers.

The system of propositions. New variables, "*p*", "*q*", etc., are introduced with a rule to the effect that any (declarative) sentence may be substituted for a variable of this kind; this includes, in addition to the sentences of the original thing language, also all general sentences with variables of any kind which may have been introduced into the language. Further, the general term "proposition" is introduced. "*p* is a proposition" may be defined by "*p* or not *p*" (or by any other sentence form yielding only analytic sentences). Therefore, every sentence of the form "...is a proposition" (where any sentence may stand in the place of the dots) is analytic. This holds, for example, for the sentence:

(*a*) "Chicago is large is a proposition".

(We disregard here the fact that the rules of English grammar require not a sentence but a that-clause as the subject of another sentence; accordingly, instead of (*a*) we should have to say "That Chicago is large is a proposition".) Predicates may be admitted whose argument expressions are sentences; these predicates may be either extensional (e.g., the customary truth-functional connectives) or not (e.g., modal predicates like "possible", "necessary", etc.). With the help of the new variables, general sentences may be formed, e.g.,

(*b*) "For every *p*, either *p* or not-*p*".

(*c*) "There is a *p* such that *p* is not necessary and not-*p* is not necessary".

(*d*) "There is a *p* such that *p* is a proposition".

(*c*) and (*d*) are internal assertions of existence. The statement "There are propositions" may be meant in the sense of (*d*); in this case it is analytic (since it follows from (*a*)) and even trivial. If, however, the statement is meant in an external sense, then it is non-cognitive.

It is important to notice that the system of rules for the linguistic expressions of the propositional framework (of which only a few rules have here been briefly indicated) is sufficient for the introduction of the framework. Any further explanations as to the nature of the propositions (i.e., the elements of the system indicated, the values of the variables "*p*", "*q*", etc.) are theoretically unnecessary because, if correct, they follow from the rules. For example, are propositions mental events (as in Russell's theory)? A look at the rules shows us that they are not, because otherwise existential statements would be of the form: "If the mental state of the person in question fulfils such and such conditions, then there is a *p* such that . . .". The fact that no references to mental conditions occur in existential statements (like (*c*), (*d*), etc.) shows that propositions are not mental entities. Further, a statement of the existence of linguistic entities (e.g., expressions, classes of expressions, etc.) must contain a reference to a language. The fact that no such reference occurs in the existential statements here shows that propositions are not linguistic entities. The fact that in these statements no reference to a subject (an observer or knower) occurs (nothing like: "There is a *p* which is necessary for Mr. *X*") shows that the propositions (and their properties, like necessity, etc.) are not subjective. Although characterizations of these or

similar kinds are, strictly speaking, unnecessary, they may nevertheless be practically useful. If they are given, they should be understood, not as ingredient parts of the system, but merely as marginal notes with the purpose of supplying to the reader helpful hints or convenient pictorial associations which may make his learning of the use of the expressions easier than the bare system of the rules would do. Such a characterization is analogous to an extra-systematic explanation which a physicist sometimes gives to the beginner. He might, for example, tell him to imagine the atoms of a gas as small balls rushing around with great speed, or the electromagnetic field and its oscillations as quasi-elastic tensions and vibrations in an ether. In fact, however, all that can accurately be said about atoms or the field is implicitly contained in the physical laws of the theories in question.[3]

The system of thing properties. The thing language contains words like "red", "hard", "stone", "house", etc., which are used for describing what things are like. Now we may introduce new variables, say "*f*", "*g*", etc., for which those words are substitutable and furthermore the general term "property". New rules are laid down which admit sentences like "Red is a property", "Red is a color", "These two pieces of paper have at least one color in common" (i.e., "There is an *f* such that *f* is a color, and . . ."). The last sentence is an internal assertion. It is of an empirical, factual nature. However, the external statement, the philosophical statement of the reality of properties – a special case of the thesis of the reality of universals – is devoid of cognitive content.

The systems of integers and rational numbers. Into a language con-

[3]In my book *Meaning and Necessity* (1947) I have developed a semantical method which takes propositions as entities designated by sentences (more specifically, as intensions of sentences). In order to facilitate the understanding of the systematic development, I added some informal, extra-systematic explanations concerning the nature of propositions. I said that the term "proposition" "is used neither for a linguistic expression nor for a subjective, mental occurrence, but rather for something objective that may or may not be exemplified in nature. . . . We apply the term 'proposition' to any entities of a certain logical type, namely, those that may be expressed by (declarative) sentences in a language" (p. 27). After some more detailed discussion concerning the relation between propositions and facts, and the nature of false propositions, I added: "It has been the purpose of the preceding remarks to facilitate the understanding of our conception of propositions. If, however, a reader should find these explanations more puzzling than clarifying, or even unacceptable, he may disregard them" (p. 31) (that is, disregard these extra-systematic explanations, not the whole theory of the propositions as intensions of sentences, as one reviewer understood). In spite of this warning, it seems that some of those readers who were puzzled by the explanations did not disregard them but thought that by raising objections against them they could refute the theory. This is analogous to the procedure of some laymen who by (correctly) criticizing the ether picture or other visualizations of physical theories thought they had refuted those theories. Perhaps the discussions in the present paper will help in clarifying the role of the system of linguistic rules for the introduction of a framework for entities on the one hand, and that of extra-systematic explanations concerning the nature of the entities on the other.

taining the framework of natural numbers we may introduce first the (positive and negative) integers as relations among natural numbers and then the rational numbers as relations among integers. This involves introducing new types of variables, expressions substitutable for them, and the general terms "integer" and "rational number".

The system of real numbers. On the basis of the rational numbers, the real numbers may be introduced as classes of a special kind (segments) of rational numbers (according to the method developed by Dedekind and Frege). Here again a new type of variables is introduced, expressions substitutable for them (e.g., "$\sqrt{2}$"), and the general term "real number".

The spatio-temporal coordinate system for physics. The new entities are the space-time points. Each is an ordered quadruple of four real numbers, called its coordinates, consisting of three spatial and one temporal coordinate. The physical state of a spatio-temporal point or region is described either with the help of qualitative predicates (e.g, "hot") or by ascribing numbers as values of a physical magnitude (e.g., mass, temperature, and the like). The step from the system of things (which does not contain space-time points but only extended objects with spatial and temporal relations between them) to the physical coordinate system is again a matter of decision. Our choice of certain features, although itself not theoretical, is suggested by theoretical knowledge, either logical or factual. For example, the choice of real numbers rather than rational numbers or integers as coordinates is not much influenced by the facts of experience but mainly due to considerations of mathematical simplicity. The restriction to rational coordinates would not be in conflict with any experimental knowledge we have, because the result of any measurement is a rational number. However, it would prevent the use of ordinary geometry (which says, e.g., that the diagonal of a square with the side 1 has the irrational value $\sqrt{2}$) and thus lead to great complications. On the other hand, the decision to use three rather than two or four spatial coordinates is strongly suggested, but still not forced upon us, by the result of common observations. If certain events allegedly observed in spiritualistic séances, e.g., a ball moving out of a sealed box, were confirmed beyond any reasonable doubt, it might seem advisable to use four spatial coordinates. Internal questions are here, in general, empirical questions to be answered by empirical investigations. On the other hand, the external questions of the reality of physical space and physical time are pseudo-questions. A question like "Are there (really) space-time points?" is ambiguous. It may be meant as an internal question; then the affirmative answer is, of course, analytic and trivial. Or it may be meant in the external sense: "Shall we introduce such and such forms into our language?"; in this case it is not a theoretical but a practical question, a

matter of decision rather than assertion, and hence the proposed formulation would be misleading. Or finally, it may be meant in the following sense: "Are our experiences such that the use of the linguistic forms in question will be expedient and fruitful?" This is a theoretical question of a factual, empirical nature. But it concerns a matter of degree; therefore a formulation in the form "real or not?" would be inadequate.

3. What does acceptance of a kind of entities mean?

Let us now summarize the essential characteristics of situations involving the introduction of a new kind of entities, characteristics which are common to the various examples outlined above.

The acceptance of a new kind of entities is represented in the language by the introduction of a framework of new forms of expressions to be used according to a new set of rules. There may be new names for particular entities of the kind in question; but some such names may already occur in the language before the introduction of the new framework. (Thus, for example, the thing language contains certainly words of the type of "blue" and "house" before the framework of properties is introduced; and it may contain words like "ten" in sentences of the form "I have ten fingers" before the framework of numbers is introduced.) The latter fact shows that the occurrence of constants of the type in question – regarded as names of entities of the new kind after the new framework is introduced – is not a sure sign of the acceptance of the new kind of entities. Therefore the introduction of such constants is not to be regarded as an essential step in the introduction of the framework. The two essential steps are rather the following. First, the introduction of a general term, a predicate of higher level, for the new kind of entities, permitting us to say of any particular entity that it belongs to this kind (e.g., "Red is a *property*", "Five is a *number*"). Second, the introduction of variables of the new type. The new entities are values of these variables; the constants (and the closed compound expressions, if any) are substitutable for the variables.[4] With the help of the variables, general sentences concerning the new entities can be formulated.

After the new forms are introduced into the language, it is possible to formulate with their help internal questions and possible answers to them. A question of this kind may be either empirical or logical; accordingly a true answer is either factually true or analytic.

[4]W. V. Quine was the first to recognize the importance of the introduction of variables as indicating the acceptance of entities. "The ontology to which one's use of language commits him comprises simply the objects that he treats as falling . . . within the range of values of his variables" (1943: 118; compare also 1939 and 1947).

From the internal questions we must clearly distinguish external questions, i.e., philosophical questions concerning the existence or reality of the total system of the new entities. Many philosophers regard a question of this kind as an ontological question which must be raised and answered *before* the introduction of the new language forms. The latter introduction, they believe, is legitimate only if it can be justified by an ontological insight supplying an affirmative answer to the question of reality. In contrast to this view, we take the position that the introduction of the new ways of speaking does not need any theoretical justification because it does not imply any assertion of reality. We may still speak (and have done so) of "the acceptance of the new entities" since this form of speech is customary; but one must keep in mind that this phrase does not mean for us anything more than acceptance of the new framework, i.e., of the new linguistic forms. Above all, it must not be interpreted as referring to an assumption, belief, or assertion of "the reality of the entities". There is no such assertion. An alleged statement of the reality of the system of entities is a pseudo-statement without cognitive content. To be sure, we have to face at this point an important question; but it is a practical, not a theoretical question; it is the question of whether or not to accept the new linguistic forms. The acceptance cannot be judged as being either true or false because it is not an assertion. It can only be judged as being more or less expedient, fruitful, conducive to the aim for which the language is intended. Judgments of this kind supply the motivation for the decision of accepting or rejecting the kind of entities.[5]

Thus it is clear that the acceptance of a linguistic framework must not be regarded as implying a metaphysical doctrine concerning the reality of the entities in question. It seems to me due to a neglect of this important distinction that some contemporary nominalists label the admission of variables of abstract types as "Platonism".[6] This is, to say the least, an

[5]For a closely related point of view on these questions see the detailed discussions in Feigl 1950: 35–62.

[6]Bernays 1935: 52–69 (reprinted in this volume). W. V. Quine, see previous footnote and a recent paper (1948). Quine does not acknowledge the distinction which I emphasize above, because according to his general conception there are no sharp boundary lines between logical and factual truth, between questions of meaning and questions of fact, between the acceptance of a language structure and the acceptance of an assertion formulated in the language. This conception, which seems to deviate considerably from customary ways of thinking, is explained in his article 1951c. When Quine classifies my logistic conception of mathematics (derived from Frege and Russell) and "platonic realism" (1948: 33), this is meant (according to a personal communication from him) not as ascribing to me agreement with Plato's metaphysical doctrine of universals, but merely as referring to the fact that I accept a language of mathematics containing variables of higher levels. With respect to the basic attitude to take in choosing a language form (an "ontology" in Quine's terminology, which seems to me misleading), there appears now to be agreement between us: "the obvious counsel is tolerance and an experimental spirit" (1948: 38).

extremely misleading terminology. It leads to the absurd consequence that the position of everybody who accepts the language of physics with its real number variables (as a language of communication, not merely as a calculus) would be called Platonistic, even if he is a strict empiricist who rejects Platonic metaphysics.

A brief historical remark may here be inserted. The non-cognitive character of the questions which we have called here external questions was recognized and emphasized already by the Vienna Circle under the leadership of Moritz Schlick, the group from which the movement of logical empiricism originated. Influenced by ideas of Ludwig Wittgenstein, the Circle rejected both the thesis of the reality of the external world and the thesis of its irreality as pseudo-statements;[7] the same was the case for both the thesis of the reality of universals (abstract entities, in our present terminology) and the nominalistic thesis that they are not real and that their alleged names are not names of anything but merely *flatus vocis*. (It is obvious that the apparent negation of a pseudo-statement must also be a pseudo-statement.) It is therefore not correct to classify the members of the Vienna Circle as nominalists, as is sometimes done. However, if we look at the basic anti-metaphysical and pro-scientific attitude of most nominalists (and the same holds for many materialists and realists in the modern sense), disregarding their occasional pseudo-theoretical formulations, then it is, of course, true to say that the Vienna Circle was much closer to those philosophers than to their opponents.

4. Abstract entities in semantics

The problem of the legitimacy and the status of abstract entities has recently again led to controversial discussions in connection with semantics. In a semantical meaning analysis certain expressions in a language are often said to designate (or name or denote or signify or refer to) certain extra-linguistic entities.[8] As long as physical things or events (e.g., Chicago or Caesar's death) are taken as designata (entities designated), no serious doubts arise. But strong objections have been raised, especially by some empiricists, against abstract entities as designata, e.g., against semantical statements of the following kind:

[7]See Carnap 1928b and Schlick 1932.

[8]See Carnap 1942, 1947. The distinction I have drawn in the latter book between the method of the name-relation and the method of intension and extension is not essential for our present discussion. The term "designation" is used in the present article in a neutral way; it may be understood as referring to the name-relation or to the intension-relation or to the extension-relation or to any similar relations used in other semantical methods.

(1) "The word 'red' designates a property of things";
(2) "The word 'color' designates a property of properties of things";
(3) "The word 'five' designates a number";
(4) "The word 'odd' designates a property of numbers";
(5) "The sentence 'Chicago is large' designates a proposition".

Those who criticize these statements do not, of course, reject the use of the expressions in question, like "red" or "five"; nor would they deny that these expressions are meaningful. But to be meaningful, they would say, is not the same as having a meaning in the sense of an entity designated. They reject the belief, which they regard as implicitly presupposed by those semantical statements, that to each expression of the types in question (adjectives like "red", numerals like "five", etc.) there is a particular real entity to which the expression stands in the relation of designation. This belief is rejected as incompatible with the basic principles of empiricism or of scientific thinking. Derogatory labels like "Platonic realism", "hypostatization", or "'Fido'-Fido principle" are attached to it. The latter is the name given by Gilbert Ryle [Meaning] to the criticized belief, which, in his view, arises by a naïve inference of analogy: just as there is an entity well known to me, viz. my dog Fido, which is designated by the name "Fido", thus there must be for every meaningful expression a particular entity to which it stands in the relation of designation or naming, i.e., the relation exemplified by "Fido"-Fido. The belief criticized is thus a case of hypostatization, i.e., of treating as names expressions which are not names. While "Fido" is a name, expressions like "red", "five", etc., are said not to be names, not to designate anything.

Our previous discussion concerning the acceptance of frameworks enables us now to clarify the situation with respect to abstract entities as designata. Let us take as an example the statement:

(*a*) "'Five' designates a number".

The formulation of this statement presupposes that our language L contains the forms of expressions which we have called the framework of numbers, in particular, numerical variables and the general term "number". If L contains these forms, the following is an analytic statement in L:

(*b*) "Five is a number".

Further, to make the statement (*a*) possible, L must contain an expression like "designates" or "is a name of" for the semantical relation of designation. If suitable rules for this term are laid down, the following is likewise analytic:

(*c*) "'Five' designates five".

(Generally speaking, any expression of the form "'...' designates ..." is an analytic statement provided the term "..." is a constant in an accepted framework. If the latter condition is not fulfilled, the expression is not a statement.) Since (*a*) follows from (*c*) and (*b*), (*a*) is likewise analytic.

Thus it is clear that *if* someone accepts the framework of numbers, then he must acknowledge (*c*) and (*b*) and hence (*a*) as true statements. Generally speaking, if someone accepts a framework for a certain kind of entities, then he is bound to admit the entities as possible designata. Thus the question of the admissibility of entities of a certain type or of abstract entities in general as designata is reduced to the question of the acceptability of the linguistic framework for those entities. Both the nominalistic critics, who refuse the status of designators or names to expressions like "red", "five", etc., because they deny the existence of abstract entities, and the skeptics, who express doubts concerning the existence and demand evidence for it, treat the question of existence as a theoretical question. They do, of course, not mean the internal question; the affirmative answer to *this* question is analytic and trivial and too obvious for doubt or denial, as we have seen. Their doubts refer rather to the system of entities itself; hence they mean the external question. They believe that only after making sure that there really is a system of entities of the kind in question are we justified in accepting the framework by incorporating the linguistic forms into our language. However, we have seen that the external question is not a theoretical question but rather the practical question whether or not to accept those linguistic forms. This acceptance is not in need of a theoretical justification (except with respect to expediency and fruitfulness), because it does not imply a belief or assertion. Ryle says that the "Fido"-Fido principle is "a grotesque theory". Grotesque or not, Ryle is wrong in calling it a theory. It is rather the practical decision to accept certain frameworks. Maybe Ryle is historically right with respect to those whom he mentions as previous representatives of the principle, viz. John Stuart Mill, Frege, and Russell. If these philosophers regarded the acceptance of a system of entities as a theory, an assertion, they were victims of the same old, metaphysical confusion. But it is certainly wrong to regard *my* semantical method as involving a belief in the reality of abstract entities, since I reject a thesis of this kind as a metaphysical pseudo-statement.

The critics of the use of abstract entities in semantics overlook the fundamental difference between the acceptance of a system of entities and an internal assertion, e.g., an assertion that there are elephants or electrons or prime numbers greater than a million. Whoever makes an internal assertion is certainly obliged to justify it by providing evidence, empirical evidence in the case of electrons, logical proof in the case of the prime numbers. The demand for a theoretical justification, correct in the case of internal assertions, is sometimes wrongly applied to the acceptance of a system of entities. Thus, for example, Ernest Nagel (1948) asks for "evidence relevant for affirming with warrant that there are such entities as infinitesimals or propositions". He characterizes the evidence required in these cases – in distinction to the empirical evidence in the case of electrons – as "in the broad sense logical and dialectical". Beyond this no hint is given as to what might be regarded as relevant evidence. Some nominalists regard the acceptance of abstract entities as a kind of superstition or myth, populating the world with fictitious or at least dubious entities, analogous to the belief in centaurs or demons. This shows again the confusion mentioned, because a superstition or myth is a false (or dubious) internal statement.

Let us take as example the natural numbers as cardinal numbers, i.e., in contexts like "Here are three books". The linguistic forms of the framework of numbers, including variables and the general term "number", are generally used in our common language of communication; and it is easy to formulate explicit rules for their use. Thus the logical characteristics of this framework are sufficiently clear (while many internal questions, i.e., arithmetical questions, are, of course, still open). In spite of this, the controversy concerning the external question of the ontological reality of the system of numbers continues. Suppose that one philosopher says: "I believe that there are numbers as real entities. This gives me the right to use the linguistic forms of the numerical framework and to make semantical statements about numbers as designata of numerals". His nominalistic opponent replies: "You are wrong; there are no numbers. The numerals may still be used as meaningful expressions. But they are not names, there are no entities designated by them. Therefore the word "number" and numerical variables must not be used (unless a way were found to introduce them as merely abbreviating devices, a way of translating them into the nominalistic thing language)." I cannot think of any possible evidence that would be regarded as relevant by both philosophers, and therefore, if actually found, would decide the controversy or at least make one of the opposite theses more probable than the other. (To construe the numbers as classes or properties of the second level, according to the Frege-Russell method, does, of course, not solve the controversy, because the first philosopher would affirm and the second deny the existence of the system of classes or properties of the second level.) Therefore I feel compelled to regard the external question

as a pseudo-question, until both parties to the controversy offer a common interpretation of the question as a cognitive question; this would involve an indication of possible evidence regarded as relevant by both sides.

There is a particular kind of misinterpretation of the acceptance of abstract entities in various fields of science and in semantics that needs to be cleared up. Certain early British empiricists (e.g., Berkeley and Hume) denied the existence of abstract entities on the ground that immediate experience presents us only with particulars, not with universals, e.g., with this red patch, but not with Redness or Color-in-General; with this scalene triangle, but not with Scalene Triangularity or Triangularity-in-General. Only entities belonging to a type of which examples were to be found within immediate experience could be accepted as ultimate constituents of reality. Thus, according to this way of thinking, the existence of abstract entities could be asserted only if one could show either that some abstract entities fall within the given, or that abstract entities can be defined in terms of the types of entity which are given. Since these empiricists found no abstract entities within the realm of sense-data, they either denied their existence, or else made a futile attempt to define universals in terms of particulars. Some contemporary philosophers, especially English philosophers following Bertrand Russell, think in basically similar terms. They emphasize a distinction between the data (that which is immediately given in consciousness, e.g., sense-data, immediately past experiences, etc.) and the constructs based on the data. Existence or reality is ascribed only to the data; the constructs are not real entities; the corresponding linguistic expressions are merely ways of speech not actually designating anything (reminiscent of the nominalists' *flatus vocis*). We shall not criticize here this general conception. (As far as it is a principle of accepting certain entities and not accepting others, leaving aside any ontological, phenomenalistic and nominalistic pseudo-statements, there cannot be any theoretical objection to it.) But if this conception leads to the view that other philosophers or scientists who accept abstract entities thereby assert or imply their occurrence as immediate data, then such a view must be rejected as a misinterpretation. References to space-time points, the electromagnetic field, or electrons in physics, to real or complex numbers and their functions in mathematics, to the excitatory potential or unconscious complexes in psychology, to an inflationary trend in economics, and the like, do not imply the assertion that entities of these kinds occur as immediate data. And the same holds for references to abstract entities as designata in semantics. Some of the criticisms by English philosophers against such references give the impression that, probably due to the misinterpretation just indicated, they accuse the semanticist not so much of bad metaphysics (as some nominalists would do) but of bad psychology. The fact that they regard a semantical method involving abstract entities not merely as doubtful and perhaps wrong, but as manifestly absurd, preposterous and grotesque, and that they show a deep horror and indignation against this method, is perhaps to be explained by a misinterpretation of the kind described. In fact, of course, the semanticist does not in the least assert or imply that the abstract entities to which he refers can be experienced as immediately given either by sensation or by a kind of rational intuition. An assertion of this kind would indeed be very dubious psychology. The psychological question as to which kinds of entities do and which do not occur as immediate data is entirely irrelevant for semantics, just as it is for physics, mathematics, economics, etc., with respect to the examples mentioned above.[9]

5. Conclusion

For those who want to develop or use semantical methods, the decisive question is not the alleged ontological question of the existence of abstract entities but rather the question whether the use of abstract linguistic forms or, in technical terms, the use of variables beyond those for things (or phenomenal data) is expedient and fruitful for the purposes for which semantical analyses are made, viz. the analysis, interpretation, clarification, or construction of languages of communication, especially languages of science. This question is here neither decided nor even discussed. It is not a question simply of yes or no, but a matter of degree. Among those philosophers who have carried out semantical analyses and thought about suitable tools for this work, beginning with Plato and Aristotle and, in a more technical way on the basis of modern logic, with C. S. Peirce and Frege, a great majority accepted abstract entities. This does, of course, not prove the case. After all, semantics in the technical sense is still in the initial phases of its development, and we must be prepared for possible fundamental changes in methods. Let us therefore admit that the nominalistic critics may possibly be right. But if so, they will have to offer better arguments than they have so far. Appeal to ontological insight will not carry much weight. The critics will have to show that it is possible to construct a semantical method which avoids all references to abstract entities and achieves by simpler means essentially the same results as the other methods.

The acceptance or rejection of abstract linguistic forms, just as the

[9]Sellars (1949: 496–504; see pp. 502f.) analyzes clearly the roots of the mistake "of taking the designation relation of semantic theory to be a reconstruction of *being present to an experience*".

acceptance or rejection of any other linguistic forms in any branch of science, will finally be decided by their efficiency as instruments, the ratio of the results achieved to the amount and complexity of the efforts required. To decree dogmatic prohibitions of certain linguistic forms instead of testing them by their success or failure in practical use is worse than futile; it is positively harmful because it may obstruct scientific progress. The history of science shows examples of such prohibitions based on prejudices deriving from religious, mythological, metaphysical, or other irrational sources, which slowed up the developments for shorter or longer periods of time. Let us learn from the lessons of history. Let us grant to those who work in any special field of investigation the freedom to use any form of expression which seems useful to them; the work in the field will sooner or later lead to the elimination of those forms which have no useful function. *Let us be cautious in making assertions and critical in examining them, but tolerant in permitting linguistic forms.*

On platonism in mathematics

PAUL BERNAYS

With your permission, I shall now address you on the subject of the present situation in research in the foundations of mathematics.

Since there remain open questions in this field, I am not in a position to paint a definitive picture of it for you. But it must be pointed out that the situation is not so critical as one could think from listening to those who speak of a foundational crisis. From certain points of view, this expression can be justified; but it could give rise to the opinion that mathematical science is shaken at its roots.

The truth is that the mathematical sciences are growing in complete security and harmony. The ideas of Dedekind, Poincaré, and Hilbert have been systematically developed with great success, without any conflict in the results.

It is only from the philosophical point of view that objections have been raised. They bear on certain ways of reasoning peculiar to analysis and set theory. These modes of reasoning were first systematically applied in giving a rigorous form to the methods of the calculus. [According to them,] the objects of a theory are viewed as elements of a totality such that one can reason as follows: For each property expressible using the notions of the theory, it is [an] objectively determinate [fact] whether there is or there is not an element of the totality which possesses this property. Similarly, it follows from this point of view that either all the elements of a set possess a given property, or there is at least one element which does not possess it.

An example of this way of setting up a theory can be found in Hilbert's axiomatization of geometry. If we compare Hilbert's axiom system to Euclid's, ignoring the fact that the Greek geometer fails to include certain [necessary] postulates, we notice that Euclid speaks of figures to be

Lecture delivered June 18, 1934, in the cycle of *Conférences internationales des Sciences mathématiques* organized by the University of Geneva, in the series on Mathematical Logic.

Translated from the French by C. D. Parsons from *L'enseignement mathématique,* 1st ser. vol. 34 (1935), pp. 52–69. Permission for the translation and inclusion of this paper in this book was kindly granted by the author and the editor of *L'enseignement mathématique.*

constructed,[1] whereas, for Hilbert, system of points, straight lines, and planes exist from the outset. Euclid postulates: One can join two points by a straight line; Hilbert states the axiom: Given any two points, there exists a straight line on which both are situated. "Exists" refers here to existence in the system of straight lines.

This example shows already that the tendency of which we are speaking consists in viewing the objects as cut off from all links with the reflecting subject.

Since this tendency asserted itself especially in the philosophy of Plato, allow me to call it "platonism."

The value of platonistically inspired mathematical conceptions is that they furnish models of abstract imagination. These stand out by their simplicity and logical strength. They form representations which extrapolate from certain regions of experience and intuition.

Nonetheless, we know that we can arithmetize the theoretical systems of geometry and physics. For this reason, we shall direct our attention to platonism in arithmetic. But I am referring to arithmetic in a very broad sense, which includes analysis and set theory.

The weakest of the "platonistic" assumptions introduced by arithmetic is that of the totality of integers. The *tertium non datur* for integers follows from it; viz.: if P is a predicate of integers, either P is true of each number, or there is at least one exception.

By the assumption mentioned, this disjunction is an immediate consequence of the logical principle of the excluded middle; in analysis it is almost continually applied.

For example, it is by means of it that one concludes that for two real numbers a and b, given by convergent series, either $a=b$ or $a<b$ or $b<a$; and likewise: a sequence of positive rational numbers either comes as close as you please to zero or there is a positive rational number less than all the members of the sequence.

At first sight, such disjunctions seem trivial, and we must be attentive in order to notice that an assumption slips in.

But analysis is not content with this modest variety of platonism; it reflects it to a stronger degree with respect to the following notions: set of numbers, sequence of numbers, and function. It abstracts from the possibility of giving definitions of sets, sequences, and functions. These notions are used in a "quasi-combinatorial" sense, by which I mean: in the sense of an analogy of the infinite to the finite.

Consider, for example, the different functions which assign to each

[1] [Translator's italics.]

member of the finite series $1, 2, \ldots, n$ a number of the same series. There are n^n functions of this sort, and each of them is obtained by n independent determinations. Passing to the infinite case, we imagine functions engendered by an infinity of independent determinations which assign to each integer an integer, and we reason about the totality of these functions.

In the same way, one views a set of integers as the result of infinitely many independent acts deciding for each number whether it should be included or excluded. We add to this the idea of the totality of these sets. Sequences of real numbers and sets of real numbers are envisaged in an analogous manner. From this point of view, constructive definitions of specific functions, sequences, and sets are only ways to pick out an object which exists independently of, and prior to, the construction.

The axiom of choice is an immediate application of the quasi-combinatorial concepts in question. It is generally employed in the theory of real numbers in the following special form. Let

$$M_1, M_2, \ldots$$

be a sequence of non-empty sets of real numbers, then there is a sequence

$$a_1, a_2, \ldots$$

such that for every index n, a_n is an element of M_n.

The principle becomes subject to objections if the effective construction of the sequence of numbers is demanded.

A similar case is that of Poincaré's impredicative definitions. An impredicative definition of a real number appeals to the hypothesis that all real numbers have a certain property P, or the hypothesis that there exists a real number with the property T.

This kind of definition depends on the assumption of [the existence of] the totality of sequences of integers, because a real number is represented by a decimal fraction, that is to say, by a special kind of sequence of integers.

It is used in particular to prove the fundamental theorem that a bounded set of real numbers always has a least upper bound.

In Cantor's theories, platonistic conceptions extend far beyond those of the theory of real numbers. This is done by iterating the use of the quasi-combinatorial concept of a function and adding methods of collection. This is the well-known method of set theory.

The platonistic conceptions of analysis and set theory have also been applied in modern theories of algebra and topology, where they have proved very fertile.

This brief summary will suffice to characterize platonism and its appli-

cation to mathematics. This application is so widespread that it is not an exaggeration to say that platonism reigns today in mathematics.

But on the other hand, we see that this tendency has been criticized in principle since its first appearance and has given rise to many discussions. This criticism was reinforced by the paradoxes discovered in set theory, even though these antinomies refute only extreme platonism.

We have set forth only a restricted platonism which does not claim to be more than, so to speak, an ideal projection of a domain of thought. But the matter has not rested there. Several mathematicians and philosophers interpret the methods of platonism in the sense of conceptual realism, postulating the existence of a world of ideal objects containing all the objects and relations of mathematics. It is this absolute platonism which has been shown untenable by the antinomies, particularly by those surrounding the Russell–Zermelo paradox.

If one hears them for the first time, these paradoxes in their purely logical form can seem to be plays on words without serious significance. Nonetheless one must consider that these abbreviated forms of the paradoxes are obtained by following out the consequences of the various requirements of absolute platonism.

The essential importance of these antinomies is to bring out the impossibility of combining the following two things: the idea of the totality of all mathematical objects and the general concepts of set and function; for the totality itself would form a domain of elements for sets, and arguments and values for functions.

We must therefore give up absolute platonism. But it must be observed that this is almost the only injunction which follows from the paradoxes. Some will think that this is regrettable, since the paradoxes are appealed to on every side. But avoiding the paradoxes does not constitute a univocal program. In particular, restricted platonism is not touched at all by the antinomies.

Still, the critique of the foundations of analysis receives new impetus from this source, and among the different possible ways of escaping from the paradoxes, eliminating platonism offered itself as the most radical.

Let us look and see how this elimination can be brought about. It is done in two steps, corresponding to the two essential assumptions introduced by platonism. The first step is to replace by constructive concepts the concepts of a set, a sequence, or a function, which I have called quasi-combinatorial. The idea of an infinity of independent determinations is rejected. One emphasizes that an infinite sequence or a decimal fraction can be given only by an arithmetical law, and one regards the continuum as a set of elements defined by such laws.

This procedure is adapted to the tendency toward a complete arithmetization of analysis. Indeed, it must be conceded that the arithmetization of analysis is not carried through to the end by the usual method. The conceptions which are applied there are not completely reducible, as we have seen, to the notion of integer and logical concepts.

Nonetheless, if we pursue the thought that each real number is defined by an arithmetical law, the idea of the totality of real numbers is no longer indispensable, and the axiom of choice is not at all evident. Also, unless we introduce auxiliary assumptions – as Russell and Whitehead do – we must do without various usual conclusions. Weyl has made these consequences very clear in his book *Das Kontinuum*.

Let us proceed to the second step of the elimination. It consists in renouncing the idea of the totality of integers. This point of view was first defended by Kronecker and then developed systematically by Brouwer.

Although several of you heard in March [1934] an authentic exposition of this method by Professor Brouwer himself, I shall allow myself a few words of explanation.

A misunderstanding about Kronecker must first be dissipated, which could arise from his often-cited aphorism that the integers were created by God, whereas everything else in mathematics is the work of man. If that were really Kronecker's opinion, he ought to admit the concept of the totality of integers.

In fact, Kronecker's method, as well as that of Brouwer, is characterized by the fact that it avoids the supposition that there exists a series of natural numbers forming a determinate ideal object.

According to Kronecker and Brouwer, one can speak of the series of numbers only in the sense of a process that is never finished, surpassing each limit which it reaches.

This point of departure carries with it the other divergences, in particular those concerning the application and interpretation of logical forms: Neither a general judgment about integers nor a judgment of existence can be interpreted as expressing a property of the series of numbers. A general theorem about numbers is to be regarded as a sort of prediction that a property will present itself for each construction of a number; and the affirmation of the existence of a number with a certain property is interpreted as an incomplete communication of a more precise proposition indicating a [particular] number having the property in question or a method for obtaining such a number; Hilbert calls it a "partial judgment."

For the same reasons the negation of a general or existential proposition about integers does not have precise sense. One must strengthen the

negation to arrive at a mathematical proposition. For example, it is to give a strengthened negation of a proposition affirming the existence of a number with a property P to say that a number with the property P cannot be given, or further, that the assumption of a number with this property leads to a contradiction. But for such strengthened negations the law of the excluded middle is no longer applicable.

The characteristic complications to be met with in Brouwer's "intuitionistic" method come from this.

For example, one may not generally make use of disjunctions like these: a series of positive terms is either convergent or divergent; two convergent sums represent either the same real number or different ones.

In the theory of integers and of algebraic numbers, we can avoid these difficulties and manage to preserve all the essential theorems and arguments.

In fact, Kronecker has already shown that the core of the theory of algebraic fields can be developed from his methodological point of view without appeal to the totality of integers.[2]

As for analysis, you know that Brouwer has developed it in accord with the requirements of intuitionism. But here one must abandon a number of the usual theorems, for example, the fundamental theorem that every continuous function has a maximum in a closed interval. Very few things in set theory remain valid in intuitionist mathematics.

We would say, roughly, that intuitionism is adapted to the theory of numbers; the semiplatonistic method, which makes use of the idea of the totality of integers but avoids quasi-combinatorial concepts, is adapted to the arithmetic theory of functions, and the usual platonism is adequate for the geometric theory of the continuum.

There is nothing astonishing about this situation, for it is a familiar procedure of the contemporary mathematician to restrict his assumptions in each domain of the science to those which are essential. By this restriction, a theory gains methodological clarity, and it is in this direction that intuitionism proves fruitful.

But as you know, intuitionism is not at all content with such a role; it opposes the usual mathematics and claims to represent the only true mathematics.

[2]To this end, Kronecker set forth in his lectures a manner of introducing the notion of algebraic number which has been almost totally forgotten, although it is the most elementary way of defining this notion. This method consists in representing algebraic numbers by the changes of sign of irreducible polynomials in one variable with rational integers as coefficients; starting from that definition, one introduces the elementary operations and relations of magnitude for algebraic numbers and proves that the ordinary laws of calculation hold; finally one shows that a polynomial with algebraic coefficients having values with different signs for two algebraic arguments a and b has a zero between a and b.

On the other hand, mathematicians generally are not at all ready to exchange the well-tested and elegant methods of analysis for more complicated methods unless there is an overriding necessity for it.

We must discuss the question more deeply. Let us try to portray more distinctly the assumptions and philosophic character of the intuitionistic method.

What Brouwer appeals to is evidence. He claims that the basic ideas of intuitionism are given to us in an evident manner by pure intuition. In relying on this, he reveals his partial agreement with Kant. But whereas for Kant there exists a pure intuition with respect to space and time, Brouwer acknowledges only the intuition of time, from which, like Kant, he derives the intuition of number.

As for this philosophic position, it seems to me that one must concede to Brouwer two essential points: first, that the concept of integer is of intuitive origin. In this respect nothing is changed by the investigations of the logicists,[3] to which I shall return later. Second, one ought not to make arithmetic and geometry correspond in the manner in which Kant did. The concept of number is more elementary than the concepts of geometry.

Still it seems a bit hasty to deny completely the existence of a geometrical intuition. But let us leave that question aside here; there are other, more urgent ones. Is it really certain that the evidence given by arithmetical intuition extends exactly as far as the boundaries of intuitionist arithmetic would require? And finally: Is it possible to draw an exact boundary between what is evident and what is only plausible?

I believe that one must answer these two questions negatively. To begin with, you know that men and even scholars do not agree about evidence in general. Also, the same man sometimes rejects suppositions which he previously regarded as evident.

An example of a much-discussed question of evidence, about which there has been controversy up to the present, is that of the axiom of parallels. I think that the criticism which has been directed against that axiom is partly explained by the special place which it has in Euclid's system. Various other axioms had been omitted, so that the parallels axiom stood out from the others by its complexity.

In this matter I shall be content to point out the following: One can have doubts concerning the evidence of geometry, holding that it extends only to topological facts or to the facts expressed by the projective axioms. One can, on the other hand, claim that geometric intuition is not exact. These opinions are self-consistent, and all have arguments in their

[3] [I have rendered '*logiciens*' throughout as 'logicists'. – Trans.]

favor. But to claim that metric geometry has an evidence restricted to the laws common to Euclidean and Bolyai-Lobachevskian geometry, an exact metrical evidence which yet would not guarantee the existence of a perfect square, seems to me rather artificial. And yet it was the point of view of a number of mathematicians.

Our concern here has been to underline the difficulties to be encountered in trying to describe the limits of evidence.

Nevertheless, these difficulties do not make it impossible that there should be anything evident beyond question, and certainly intuitionism offers some such. But does it confine itself completely within the region of this elementary evidence? This is not completely indubitable, for the following reason: Intuitionism makes no allowance for the possibility that, for very large numbers, the operations required by the recursive method of constructing numbers can cease to have a concrete meaning. From two integers k, l one passes immediately to k^l; this process leads in a few steps to numbers which are far larger than any occurring in experience, e.g., $67^{(257^{729})}$.

Intuitionism, like ordinary mathematics, claims that this number can be represented by an Arabic numeral. Could not one press further the criticism which intuitionism makes of existential assertions and raise the question: What does it mean to claim the existence of an Arabic numeral for the foregoing number, since in practice we are not in a position to obtain it?

Brouwer appeals to intuition, but one can doubt that the evidence for it really is intuitive. Isn't this rather an application of the general method of analogy, consisting in extending to inaccessible numbers the relations which we can concretely verify for accessible numbers? As a matter of fact, the reason for applying this analogy is strengthened by the fact that there is no precise boundary between the numbers which are accessible and those which are not. One could introduce the notion of a "practicable" procedure, and implicitly restrict the import of recursive definitions to practicable operations. To avoid contradictions, it would suffice to abstain from applying the principle of the excluded middle to the notion of practicability. But such abstention goes without saying for intuitionism.

I hope I shall not be misunderstood: I am far from recommending that arithmetic be done with this restriction. I am concerned only to show that intuitionism takes as its basis propositions which one can doubt and in principle do without, although the resulting theory would be rather meager.

It is therefore not absolutely indubitable that the domain of complete evidence extends to all of intuitionism. On the other hand, several mathematicians recognize the complete evidence of intuitionistic arithmetic and moreover maintain that the concept of the series of numbers is evident in the following sense: The affirmation of the existence of a number does not require that one must, directly or recursively, give a bound for this number. Besides, we have just seen how far beyond a really concrete presentation such a limitation would be.

In short, the point of view of intuitive evidence does not decide uniquely in favor of intuitionism.

In addition, one must observe that the evidence which intuitionism uses in its arguments is not always of an immediate character. Abstract reflections are also included. In fact, intuitionists often use statements, containing a general hypothesis, of the form 'If every number n has the property $A(n)$, then B holds.'

Such a statement is interpreted intuitionistically in the following manner: 'If it is proved that every number n possesses the property $A(n)$, then B.' Here we have a hypothesis of an abstract kind, because since the methods of demonstration are not fixed in intuitionism, the condition that something is proved is not intuitively determined.

It is true that one can also interpret the given statement by viewing it as a partial judgment, i.e., as the claim that there exists a proof of B from the given hypothesis, a proof which would be effectively given.[4] (This is approximately the sense of Kolmogorov's interpretation of intuitionism.) In any case, the argument must start from the general hypothesis, which cannot be intuitively fixed. It is therefore an abstract reflection.

In the example just considered, the abstract part is rather limited. The abstract character becomes more pronounced if one superposes hypotheses; i.e., when one formulates propositions like the following: 'If from the hypothesis that $A(n)$ is valid for every n, one can infer B, then C holds,' or 'If from the hypothesis that A leads to a contradiction, a contradiction follows, then B,' or briefly 'If the absurdity of A is absurd, then B.' This abstractness of statements can be still further increased.

It is by the systematic application of these forms of abstract reasoning that Brouwer has gone beyond Kronecker's methods and succeeded in establishing a general intuitionistic logic, which has been systematized by Heyting.

If we consider this intuitionistic logic, in which the notions of consequence are applied without reservation, and we compare the method used here with the usual one, we notice that the characteristic general

[4] ["... c'est à dire comme une indication d'un raisonnement conduisant de la dite hypothèse à la conclusion B, raisonnement qu'on présente effectivement."]

feature of intuitionism is not that of being founded on pure intuition, but rather [that of being founded] on the relation of the reflecting and acting subject to the whole development of science.

This is an extreme methodological position. It is contrary to the customary manner of doing mathematics, which consists in establishing theories detached as much as possible from the thinking subject.

This realization leads us to doubt that intuitionism is the sole legitimate method of mathematical reasoning. For even if we admit that the tendency away from the [thinking] subject has been pressed too far under the reign of platonism, this does not lead us to believe that the truth lies in the opposite extreme. Keeping both possibilities in mind, we shall rather aim to bring about in each branch of science an adaptation of method to the character of the object investigated.

For example, for number theory the use of the intuitive concept of a number is the most natural. In fact, one can thus establish the theory of numbers without introducing an axiom, such as that of complete induction, or axioms of infinity like those of Dedekind and Russell.

Moreover, in order to avoid the intuitive concept of number, one is led to introduce a more general concept, like that of a proposition, a function, or an arbitrary correspondence, concepts which are in general not objectively defined. It is true that such a concept can be made more definite by the axiomatic method, as in axiomatic set theory, but then the system of axioms is quite complicated.

You know that Frege tried to deduce arithmetic from pure logic by viewing the latter as the general theory of the universe of mathematical objects. Although the foundation of this absolutely platonistic enterprise was undermined by the Russell–Zermelo paradox, the school of logicists has not given up the idea of incorporating arithmetic in a system of logic. In place of absolute platonism, they have introduced some initial assumptions. But because of these, the system loses the character of pure logic.

In the system of *Principia Mathematica,* it is not only the axioms of infinity and reducibility which go beyond pure logic, but also the initial conception of a universal domain of individuals and of a domain of predicates. It is really an *ad hoc* assumption to suppose that we have before us the universe of things divided into subjects and predicates, ready-made for theoretical treatment.

But even with such auxiliary assumptions, one cannot successfully incorporate the whole of arithmetic into the system of logic. For, since this system is developed according to fixed rules, one would have to be able to obtain by means of a fixed series of rules all the theorems of arithmetic. But this is not the case; as Gödel has shown, arithmetic goes beyond each given formalism. (In fact, the same is true of axiomatic set theory.)

Besides, the desire to deduce arithmetic from logic derives from the traditional opinion that the relation of logic to arithmetic is that of general to particular. The truth, it seems to me, is that mathematical abstraction does not have a lesser degree than logical abstraction, but rather another direction.

These considerations do not detract at all from the intrinsic value of that research of logicists which aims at developing logic systematically and formalizing mathematical proofs. We were concerned here only with defending the thesis that for the theory of numbers, the intuitive method is more suitable.

On the other hand, for the theory of the continuum, given by analysis, the intuitionist method seems rather artificial. The idea of the continuum is a geometrical idea which analysis expresses in terms of arithmetic.

Is the intuitionist method of representing the continuum better adapted to the idea of the continuum than the usual one?

Weyl would have us believe this. He reproaches ordinary analysis for decomposing the continuum into single points. But isn't this reproach better addressed to semiplatonism, which views the continuum as a set of arithmetical laws? The fact is that for the usual method there is a completely satisfying analogy between the manner in which a particular point stands out from the continuum and the manner in which a real number defined by an arithmetical law stands out from the set of all real numbers, whose elements are in general only implicitly involved, by virtue of the quasi-combinatorial concept of a sequence.

This analogy seems to me to agree better with the nature of the continuum than that which intuitionism establishes between the fuzzy character of the continuum and the uncertainties arising from unsolved arithmetical problems.

It is true that in the usual analysis the notion of a continuous function, and also that of a differentiable function, have a generality going far beyond our intuitive representation of a curve. Nevertheless, in this analysis we can establish the theorem of the maximum of a continuous function and Rolle's theorem, thus rejoining the intuitive conception.

Intuitionist analysis, even though it begins with a much more restricted notion of a function, does not arrive at such simple theorems; they must instead be replaced by more complex ones. This stems from the fact that on the intuitionistic conception, the continuum does not have the character of a totality, which undeniably belongs to the geometric idea of the

continuum. And it is this characteristic of the continuum which would resist perfect arithmetization.

These considerations lead us to notice that the duality of arithmetic and geometry is not unrelated to the opposition between intuitionism and platonism. The concept of number appears in arithmetic. It is of intuitive origin, but then the idea of the totality of numbers is superimposed. On the other hand, in geometry the platonistic idea of space is primordial, and it is against this background that the intuitionist procedures of constructing figures take place.

This suffices to show that the two tendencies, intuitionist and platonist, are both necessary; they complement each other, and it would be doing oneself violence to renounce one or the other.

But the duality of these two tendencies, like that of arithmetic and geometry, is not a perfect symmetry. As we have noted, it is not proper to make arithmetic and geometry correspond completely: the idea of number is more immediate to the mind than the idea of space. Likewise, we must recognize that the assumptions of platonism have a transcendent character which is not found in intuitionism.

It is also this transcendent character which requires us to take certain precautions in regard to each platonistic assumption. For even when such a supposition is not at all arbitrary and presents itself naturally to the mind, it can still be that the principle from which it proceeds permits only a restricted application, outside of which one would fall into contradiction.

We must be all the more careful in the face of this possibility, since the drive for simplicity leads us to make our principles as broad as possible. And the need for a restriction is often not noticed.

This was the case, as we have seen, for the principle of totality, which was pressed too far by absolute platonism. Here it was only the discovery of the Russell–Zermelo paradox which showed that a restriction was necessary.

Thus it is desirable to find a method to make sure that the platonistic assumptions on which mathematics is based do not go beyond permissible limits. The assumptions in question reduce to various forms of the principle of totality and of the principle of analogy or of the permanence of laws. And the condition restricting the application of these principles is none other than that of the consistency of the consequences which are deduced from the fundamental assumptions.

As you know, Hilbert is trying to find ways of giving us such assurances of consistency, and his proof theory has this as its goal.

This theory relies in part on the results of the logicists. They have shown that the arguments applied in arithmetic, analysis, and set theory can be formalized. That is, they can be expressed in symbols and as symbolic processes which unfold according to fixed rules. To primitive propositions correspond initial formulae, and to each logical deduction corresponds a sequence of formulae derivable from one another according to given rules. In this formalism, a platonistic assumption is represented by an initial formula or by a rule establishing a way of passing from formulae already obtained to others. In this way, the investigation of the possibilities of proof reduces to problems like those which are found in elementary number theory. In particular, the consistency of the theory will be proved if one succeeds in proving that it is impossible to deduce two mutually contradictory formulae A and $\bar{A}$ (with the bar representing negation). This statement which is to be proved is of the same structure as that, for example, asserting the impossibility of satisfying the equation $a^2=2b^2$ by two integers a and b.

Thus by symbolic reduction, the question of the consistency of a theory reduces to a problem of an elementary arithmetical character.

Starting from this fundamental idea, Hilbert has sketched a detailed program of a theory of proof, indicating the leading ideas of the arguments (for the main consistency proofs). His intention was to confine himself to intuitive and combinatorial considerations; his "finitary point of view" was restricted to these methods.

In this framework, the theory was developed up to a certain point. Several mathematicians have contributed to it: Ackermann, von Neumann, Skolem, Herbrand, Gödel, Gentzen.

Nonetheless, these investigations have remained within a relatively restricted domain. In fact, they did not even reach a proof of the consistency of the axiomatic theory of integers. It is known that the symbolic representation of this theory is obtained by adding to the ordinary logical calculus formalizations of Peano's axioms and the recursive definitions of sum $(a+b)$ and product $(a \cdot b)$.

Light was shed on this situation by a general theorem of Gödel, according to which a proof of the consistency of a formalized theory cannot be represented by means of the formalism considered. From this theorem, the following more special proposition follows: It is impossible to prove by elementary combinatorial methods the consistency of a formalized theory which can express every elementary combinatorial proof of an arithmetical proposition.

Now it seems that this proposition applies to the formalism of the axiomatic theory of numbers. At least, no attempt made up to now has given us any example of an elementary combinatorial proof which cannot be expressed in this formalism, and the methods by which one can, in

the cases considered, translate a proof into the aforementioned formalism seem to suffice in general.

Assuming that this is so,[5] we arrive at the conclusion that means more powerful than elementary combinatorial methods are necessary to prove the consistency of the axiomatic theory of numbers. A new discovery of Gödel and Gentzen leads us to such a more powerful method. They have shown (independently of one another) that the consistency of intuitionist arithmetic implies the consistency of the axiomatic theory of numbers. This result was obtained by using Heyting's formalization of intuitionist arithmetic and logic. The argument is conducted by elementary methods, in a rather simple manner. In order to conclude from this result that the axiomatic theory of numbers is consistent, it suffices to assume the consistency of intuitionist arithmetic.

This proof of the consistency of axiomatic number theory shows us, among other things, that intuitionism, by its abstract arguments, goes essentially beyond elementary combinatorial methods.

The question which now arises is whether the strengthening of the method of proof theory obtained by admitting the abstract arguments of intuitionism would put us into a position to prove the consistency of analysis. The answer would be very important and even decisive for proof theory, and even, it seems to me, for the role which is to be attributed to intuitionistic methods.

Research in the foundations of mathematics is still developing. Several basic questions are open, and we do not know what we shall discover in this domain. but these investigations excite our curiosity by their changing perspectives, and that is a sentiment which is not aroused to the same degree by the more classical parts of science, which have attained greater perfection.

I wish to thank Professor Wavre, who was kind enough to help me improve the text of this lecture for publication. I also thank M. Rueff, who was good enough to look over the first draft to improve the French.

[5]In trying to demonstrate the possibility of translating each elementary combinatorial proof of an arithmetical proposition into the formalism of the axiomatic theory of numbers, we are confronted with the difficulty of delimiting precisely the domain of elementary combinatorial methods.

What numbers could not be

PAUL BENACERRAF

The attention of the mathematician focuses primarily upon mathematical structure, and his intellectual delight arises (in part) from seeing that a given theory exhibits such and such a structure, from seeing how one structure is "modelled" in another, or in exhibiting some new structure and showing how it relates to previously studied ones.... But...the mathematician is satisfied so long as he has some "entities" or "objects" (or "sets" or "numbers" or "functions" or "spaces" or "points") to work with, and he does not inquire into their inner character or ontological status.

The philosophical logician, on the other hand, is more sensitive to matters of ontology and will be especially interested in the kind or kinds of entities there are actually.... He will not be satisfied with being told merely that such and such entities exhibit such and such a mathematical structure. He will wish to inquire more deeply into what these entities are, how they relate to other entities.... Also he will wish to ask whether the entity dealt with is *sui generis* or whether it is in some sense *reducible* to (or *constructible* in terms of) other, perhaps more fundamental entities.

– R. M. MARTIN, *Intension and Decision*

We can...by using...[our]...definitions say what is meant by
"the number 1 + 1 belongs to the concept F"
and then, using this, give the sense of the expression
"the number 1 + 1 + 1 belongs to the concept F"
and so on; but we can never...decide by means of our definitions whether any concept has the number Julius Caesar belonging to it, or whether that same familiar conqueror of Gaul is a number or not.

– G. FREGE, *The Foundations of Arithmetic*

I. The education

Imagine Ernie and Johnny, sons of two militant logicists – children who have not been taught in the vulgar (old-fashioned) way but for whom the

Much of the work on this paper was done while the author held a Procter and Gamble Faculty Fellowship at Princeton University. This is gratefully acknowledged. I am indebted

pedagogical order of things has been the epistemological order. They did not learn straight off how to count. Instead of beginning their mathematical training with arithmetic as ordinary men know it, they first learned logic – in their case, actually set theory. Then they were told about the numbers. But to tell people in their position about the numbers was an easy task – very much like the one which faced Monsieur Jourdain's tutor (who, oddly enough, was a philosopher). The parents of our imagined children needed only to point out what aspect or part of what the children already knew, under other names, was what ordinary people called "numbers." Learning the numbers merely involved learning new names for familiar sets. Old (set-theoretic) truths took on new (number-theoretic) clothing.

The way in which this was done will, however, bear some scrutiny and re-examination. To facilitate the exposition, I will concentrate on Ernie and follow his arithmetical education to its completion. I will then return to Johnny.

It might have gone as follows. Ernie was told that there was a set whose members were what ordinary people referred to as the (natural) numbers, and that these were what he had known all along as the elements of the (infinite) set $\mathcal{N}$. He was further told that there was a relation defined on these "numbers" (henceforth I shall usually omit the shudder quotes), the *less-than* relation, under which the numbers were well ordered. This relation, he learned, was really the one, defined on $\mathcal{N}$, for which he had always used the letter "R." And indeed, speaking intuitively now, Ernie could verify that every nonempty subset of $\mathcal{N}$ contained a "least" element – that is, one that bore R to every other member of the subset. He could also show that nothing bore R to itself, and that R was transitive, antisymmetric, irreflexive, and connected in $\mathcal{N}$. In short, the elements of $\mathcal{N}$ formed a progression, or series, under R.

And then there was 1, the smallest number (for reasons of future convenience we are ignoring 0). Ernie learned that what people had been referring to as 1 was really the element a of $\mathcal{N}$, the first, or least, element of $\mathcal{N}$ under R. Talk about "successors" (each number is said to have one) was easily translated in terms of the concept of the "next" member of $\mathcal{N}$ (under R). At this point, it was no trick to show that the assumptions made by ordinary mortals about numbers were in fact theorems for Ernie. For on the basis of his theory, he could establish Peano's axioms – an advantage which he enjoyed over ordinary mortals, who must more or

to Paul Ziff for his helpful comments on an earlier draft of this paper.

Reprinted with the kind permission of the editors from the *Philosophical Review* 74(1965): 47–73.

less take them as given, or self-evident, or meaningless-but-useful, or what have you.[1]

There are two more things that Ernie had to learn before he could truly be said to be able to speak with the vulgar. It had to be pointed out to him which operations on the members of $\mathfrak{N}$ were the ones referred to as "addition," "multiplication," "exponentiation," and so forth. And here again he was in a position of epistemological superiority. For whereas ordinary folk had to introduce such operations by recursive definition, a euphemism for postulation, he was in a position to show that these operations could be *explicity* defined. So the additional postulates assumed by the number people were also shown to be derivable in his theory, once it was seen which set-theoretic operations addition, multiplication, and so forth really are.

The last element needed to complete Ernie's education was the explanation of the *applications* of these devices: counting and measurement. For they employ concepts beyond those as yet introduced. But fortunately, Ernie was in a position to see what it was that he was doing that corresponded to these activities (we will concentrate on counting, assuming that measurement can be explained either similarly or in terms of counting).

There are two kinds of counting, corresponding to transitive and intransitive uses of the verb "to count." In one, "counting" admits of a direct object, as in "counting the marbles"; in the other it does not. The case I have in mind is that of the preoperative patient being prepared for the operating room. The ether mask is placed over his face and he is told to count, as far as he can. He has not been instructed to count anything at all. He has merely been told to count. A likely story is that we normally learn the first few numbers in connection with sets having that number of members – that is, in terms of *transitive* counting (thereby learning the use of numbers) and then learn how to generate "the rest" of the numbers. Actually, "the rest" always remains a relatively vague matter. Most of us simply learn that we will never run out, that our notation will extend as far as we will ever need to count. Learning these words, and how to repeat them in the right order, is learning *intransitive* counting. Learning their use as measures of sets is learning *transitive* counting. Whether we learn one kind of counting before the other is immaterial so far as the initial numbers are concerned. What is certain, and not immaterial, is that we will have to learn some recursive procedure for generating the *notation* in the proper order before we have learned to count transitively, for to do the latter is, either directly or

[1]The details of the proofs need not detain us.

indirectly, to correlate the elements of the number series with the members of the set we are counting. It seems, therefore, that it is possible for someone to learn to count intransitively without learning to count transitively. But not vice versa. This is, I think, a mildly significant point. But what *is* transitive counting, exactly?

To count the members of a set is to determine the cardinality of the set. It is to establish that a particular relation C obtains between the set and one of the numbers – that is, one of the elements of $\mathfrak{N}$ (we will restrict ourselves to counting finite sets here). Practically speaking, and in simple cases, one determines that a set has k elements by taking (sometimes metaphorically) its elements one by one as we say the numbers one by one (starting with 1 and in order of magnitude, the last number we say being k). To count the elements of some k-membered set b is to establish a one-to-one correspondence between the elements of b and the elements of $\mathfrak{N}$ less than or equal to k. The relation "pointing-to-each-member-of-b-in-turn-while-saying-the-numbers-up-to-and-including-k" establishes such a correspondence.

Since Ernie has at his disposal the machinery necessary to show of any two equivalent finite sets that such a correspondence exists between them, it will be a theorem of his system that any set has k members if and only if it can be put into one-to-one correspondence with the set of numbers less than or equal to k.[2]

Before Ernie's education (and the analysis of number) can be said to have been completed, one last condition on R should be mentioned: that R must be at least recursive, and possibly even primitive recursive. I have never seen this condition included in the analysis of number, but it seems to me so obviously required that its inclusion is hardly debatable. We

[2]It is not universally agreed that these last two parts of our account (defining the operations and defining cardinality) are indeed required for an adequate explication of number. W. V. Quine, for one, explicitly denies that anything need be done other than provide a progression to serve as the numbers. He states: "The condition upon all acceptable explications of number . . . can be put . . . : *any progression* – i.e., any infinite series each of whose members has only finitely many precursors – will do nicely. Russell once held that a further condition had to be met, to the effect that there be a way of applying one's would-be numbers to the measurement of multiplicity: a way of saying that (1) There are n objects x such that Fx. This, however, was a mistake, for any progression can be fitted to that further condition. For (1) can be paraphrased as saying that the numbers less than n [Quine uses 0 as well] admit of correlation with the objects x such that Fx. This requires that our apparatus include enough of the elementary theory of relations for talk of correlation, or one–one relation; but it requires nothing special about numbers except that they form a progression" (Quine 1960: 262–3). I would disagree. The explanation of cardinality – i.e., of the use of numbers for "transitive counting," as I have called it – is part and parcel of the explication of number. Indeed, if it may be excluded on the grounds Quine offers, we might as well say that there are *no* necessary conditions, since the only one he cites is hardly necessary, provided "that our apparatus contain enough of the theory of sets to contain a progression." But I will return to this point.

have already seen that Quine denies (by implication) that this constitutes an additional requirement: "The condition upon all acceptable explications of number...can be put...: any *progression* – i.e., any infinite series each of whose members has only finitely many precursors – will do nicely" (see note 3). But suppose, for example, that one chose the progression $A = a_1, a_2, a_3, \ldots a_n, \ldots$ obtained as follows. Divide the positive integers into two sequences B and C, within each sequence letting the elements come in order of magnitude. Let B (that is, $b_1, b_2, \ldots$) be the sequence of Gödel numbers of valid formulas of quantification theory, under some suitable numbering, and let C (that is, $c_1, c_2 \ldots$) be the sequence of positive integers which are not numbers of valid formulas of quantification theory under that numbering (in order of magnitude in each case). Now in the sequence A, for each n let $a_{2n-1} = b_n$ and $a_{2n} = c_n$. Clearly A, though a progression, is not recursive, much less primitive recursive. Just as clearly, this progression would be unusable as the numbers – and the reason is that we expect that if we know which numbers two expressions designate, we are able to calculate in a finite number of steps which is the "greater" (in this case, which one comes later in A).[3] More dramatically, if told that set b has n members, and that c has m, it should be possible to determine in a finite number of steps which has more members. Yet it is precisely that which is not possible here. This ability (to tell in a finite number of steps which of two numbers is the greater) is connected with (both transitive and intransitive) counting, since its possibility is equivalent to the possibility of generating ("saying") the numbers in order of magnitude (that is, in their order in A). You could not know that you were saying them in order of magnitude since, no recursive rule existing for generating its members, you could not know what their order of magnitude should be. This is, of course, a very strong claim. There are two questions here, both of which are interesting and neither of which could conceivably receive discussion in this paper. (1) Could a human being be a decision procedure for non-

[3]There is, of course, a difficulty with the notion of "knowing which numbers two expressions designate." It is the old one illustrated by the following example. Abraham thinks of a number, and so does Isaac. Call Abraham's number a and Isaac's i. Is a greater than i? I know which number a refers to: Abraham's. And similarly with i. But that brings me no closer to deciding which is the greater. This can be avoided, however, by requiring that numbers be given in canonical notation, as follows. Let the usual (recursive) definition of the numbers serve to define the *set* of "numbers," but not to establish their order. Then take the above definition of A as defining the *less-than* relation among the members of that set, thus defining the *progression*. (The fact that the nonrecursive progression that I use is a progression of *numbers* is clearly inessential to the point at issue. I use it here merely to avoid the elaborate circumlocutions that would result from doing everything set-theoretically. One could get the same effect by letting the "numbers" be formulas of quantification theory, instead of their Gödel numbers, and using the formulas autonymously.)

recursive sets, or is the human organism at best a Turing machine (in the relevant respect)? If the latter, then there could not exist a human being who could generate the sequence A, much less *know* that this is what he was doing. Even if the answer to (1), however, is that a human being *could* be (act or be used as) such a decision procedure, the following question would still arise and need an answer: (2) could he *know* all truths of the form $i<j$ (in A)? And it seems that what constitutes knowledge might preclude such a possibility.

But I have digressed enough on this issue. The main point is that the "$<$" relation over the numbers must be recursive. Obviously I cannot give a rigorous proof that this is a requirement, because I cannot prove that man is at best a Turing machine. That the requirement is met by the usual "$<$" relation among numbers – the paradigm of a primitive recursive relation – and has also been met in every detailed analysis ever proposed constitutes good evidence for its correctness.[4] I am just making explicit what almost everyone takes for granted. Later in this paper, we will see that one plausible account of why this is taken for granted connects very closely with one of the views I will be urging.

So it was thus that Ernie learned that he had really been doing number theory all his life (I guess in much the same way that *our* children will learn this surprising fact about themselves if the *nouvelle vague* of mathematics teachers manages to drown them all).

It should be clear that Ernie's education is now complete. He has learned to speak with the vulgar, and it should be obvious to all that my earlier description was correct. He had at his disposal all that was needed for the concept of number. One might even say that he already possessed the concepts of number, cardinal, ordinal, and the usual operations on them, and needed only to learn a different vocabulary. It is my claim that there is nothing having to do with the task of "reducing" the concept of number to logic (or set theory) that has not been done above, or that could not be done along the lines already marked out.

To recapitulate: It was necessary (1) to give definitions of "1," "number," and "successor," and "+," "×," and so forth, on the basis of which the laws of arithmetic could be derived; and (2) to explain the "extramathematical" uses of numbers, the principal one being counting – thereby introducing the concept of *cardinality* and cardinal number.

I trust that both were done satisfactorily, that the preceding contains all the elements of a correct account, albeit somewhat incompletely.

[4]Needless to say, it is trivially met in any analysis that provides an effective correlation between the names of the "numbers" of the analysis and the more common names under which we know those numbers.

None of the above was essentially new; I apologize for the tedium of expounding these details yet another time, but it will be crucial to my point that the sufficiency of the above account be clearly seen. For if it is sufficient, presumably Ernie *now* knows which sets the numbers are.

II. The dilemma

The story told in the previous section could have been told about Ernie's friend Johnny as well. For his education also satisfied the conditions just mentioned. Delighted with what they had learned, they started proving theorems about numbers. Comparing notes, they soon became aware that something was wrong, for a dispute immediately ensued about whether or not 3 belonged to 17. Ernie said that it did, Johnny that it did not. Attempts to settle this by asking ordinary folk (who had been dealing with numbers *as* numbers for a long time) understandably brought only blank stares. In support of his view, Ernie pointed to his theorem that for any two numbers, x and y, x is less than y if and only if x belongs to y and x is a proper subset of y. Since by common admission 3 is less than 17 it followed that 3 belonged to 17. Johnny, on the other hand, countered that Ernie's "theorem" was mistaken, for given two numbers, x and y, x belongs to y if and only if y is the successor of x. These were clearly incompatible "theorems." Excluding the possibility of the inconsistency of their common set theory, the incompatibility must reside in the definitions. First "less-than." But both held that x is less than y if and only if x bears R to y. A little probing, however, revealed the source of the trouble. For Ernie, the successor under R of a number x was the set consisting of x and all the members of x, while for Johnny the successor of x was simply $[x]$, the unit set of x – the set whose only member is x. Since for each of them 1 was the unit set of the null set, their respective progressions were

(i) $[\emptyset], [\emptyset,[\emptyset]], [\emptyset,[\emptyset],[\emptyset,[\emptyset]]], \ldots$ for Ernie

and

(ii) $[\emptyset], [[\emptyset]], [[[\emptyset]]], \ldots$ for Johnny.

There were further disagreements. As you will recall, Ernie had been able to prove that a set had n members if and only if it could be put into one-to-one correspondence with the set of numbers less than or equal to n. Johnny concurred. But they disagreed when Ernie claimed further that a set had n members if and only if it could be put into one-to-one correspondence with the number n itself. For Johnny, every number is

single-membered. In short, their cardinality relations were different. For Ernie, 17 had 17 members, while for Johnny it had only one.[5] And so it went.

Under the circumstances, it became perfectly obvious why these disagreements arose. But what did not become perfectly obvious was how they were to be resolved. For the problem was this:

If the conclusions of the previous section are correct, then both boys have been given correct accounts of the numbers. Each was told by his father which set the set of numbers really was. Each was taught which object – whose independent existence he was able to prove – was the number 3. Each was given an account of the meaning (and reference) of number words to which no exception could be taken and on the basis of which all that we know about or do with numbers could be explained. Each was taught that some particular set of objects contained what people who used number words were really referring to. But the sets were different in each case. And so were the relations defined on these sets – including crucial ones, like cardinality and the like. But if, as I think we agreed, the account of the previous section was correct – not only as far as it went but correct in that it contained conditions which were both necessary and *sufficient* for any correct account of the phenomena under discussion, then the fact that they disagree on which particular sets the numbers are is fatal to the view that each number is some particular set. For if the number 3 is in fact some particular set b, it cannot be that two correct accounts of the meaning of "3" – and therefore also of its reference – assign two different sets to 3. For if it is true that for some set b, $3=b$, then it cannot be true that for some set c, different from b, $3=c$. But if Ernie's account is adequate in virtue of satisfying the conditions spelled out in Section I, so is Johnny's, for it too satisfies those conditions. We are left in a quandary. We have two (infinitely many, really) accounts of the meaning of certain words ("number," "one," "seventeen," and so forth) each of which satisfies what appear to be necessary and sufficient conditions for a correct account. Although there are differences between the two accounts, it appears that both are correct in virtue of satisfying common conditions. If so, the differences are incidental and do not affect correctness. Furthermore, in Fregean terminology, each account fixes the *sense* of the words whose analysis it provides. Each account must also, therefore, fix the *reference* of these

[5]Some of their type-theoretical cousins had even more peculiar views – for to be of cardinality 5 a set had to *belong* to one of the numbers 5. I say "some of" because others did not use that definition of cardinality, or of numbers, but sided either with Ernie or with Johnny.

expressions. Yet, as we have seen, one way in which these accounts differ is in the referents assigned to the terms under analysis. This leaves us with the following alternatives:

(A) Both are right in their contentions: each account contained conditions each of which was necessary and which were jointly sufficient. Therefore 3=[[[Ø]]], and 3=[Ø,[Ø],[Ø]]].

(B) It is not the case that both accounts were correct; that is at least one contained conditions which were not necessary and possibly failed to contain further conditions which, taken together with those remaining, would make a set of sufficient conditions.

(A) is, of course, absurd. So we must explore (B).

The two accounts agree in over-all structure. They disagree when it comes to fixing the referents for the terms in question. Given the identification of the numbers as some particular set of sets, the two accounts generally agree on the relations defined on that set; under both, we have what is demonstrably a recursive progression and a successor function which follows the order of that progression. Furthermore, the notions of cardinality are defined in terms of the progression, insuring that it becomes a theorem for each n that a set has n members if and only if it can be put into one-to-one correspondence with the set of numbers less than or equal to n. Finally, the ordinary arithmetical operations are defined for these "numbers." They do differ in the way in which cardinality is defined, for in Ernie's account the fact that the number n had n members was exploited to define the notion of having n members. In all other respects, however, they agree.

Therefore, if it is not the case that both 3=[[[Ø]]] and 3=[Ø],[Ø,[Ø]]], which it surely is not, then at least one of the corresponding accounts is incorrect as a result of containing a condition that is not necessary. It may be incorrect in other respects as well, but at least that much is clear. I can distinguish two possibilities again: either all the conditions just listed, which both of these accounts share, are necessary for a correct and complete account, or some are not. Let us assume that the former is the case, although I reserve the right to discard this assumption if it becomes necessary to question it.

If all the conditions they share are necessary, then the superfluous conditions are to be found among those that are not shared. Again there are two possibilities: either at least one of the accounts satisfying the conditions we are assuming to be necessary, but which assigns a definite set to each number, is correct, or none are. Clearly no two different ones can be, since they are not even extensionally equivalent, much less intensionally. Therefore exactly one is correct or none is. But then the correct

one must be the one that picks out which set of sets is *in fact* the numbers. We are now faced with a crucial problem: if there exists such a "correct" account, do there also exist arguments which will show it to be the correct one? Or does there exist a particular set of sets b, which is *really* the numbers, but such that there exists no argument one can give to establish that it, and not, say, Ernie's set $\mathfrak{N}$, is really the numbers? It seems altogether too obvious that this latter possibility borders on the absurd. If the numbers constitute one particular set of sets, and not another, then there must be arguments to indicate which. In urging this I am not committing myself to the decidability by proof of every mathematical question – for I consider this neither a mathematical question nor one amenable to proof. The answer to the question I am raising will follow from an analysis of questions of the form "Is $n = \ldots$?" It will suffice for now to point to the difference between our question and

> Is there a greatest prime p such that $p+2$ is also prime?

or even

> Does there exist an infinite set of real numbers equivalent with neither the set of integers nor with the set of all real numbers?

In awaiting enlightenment on the true identity of 3 we are not awaiting a proof of some deep theorem. Having gotten as far as we have without settling the identity of 3, we can go no further. We do not know what a proof of that *could* look like. The notion of "correct account" is breaking loose from its moorings if we admit of the possible existence of unjustifiable but correct answers to questions such as this. To take seriously the question "Is 3=[[[Ø]]]?" *tout court* (and not elliptically for "in Ernie's account?"), in the absence of any way of settling it, is to lose one's bearings completely. No, if such a question has an answer, there are arguments supporting it, and if there are no such arguments, then there *is* no "correct" account that discriminates among all the accounts satisfying the conditions of which we reminded ourselves a couple of pages back.

How then might one distinguish *the* correct account from all the possible ones? Is there a set of sets that has a greater claim to be the numbers than any other? Are there reasons one can offer to single out that set? Frege chose as the number 3 the extension of the concept "equivalent with some 3-membered set"; that is, for Frege a number was an equivalence class – the class of all classes equivalent with a given class. Although an appealing notion, there seems little to recommend it over, say, Ernie's. It has been argued that this is a more fitting account because number words are really class predicates, and that this account reveals that fact. The view is that in saying that there are *n F*'s you are predicating *n*-hood of *F*, just as in saying that red is a color you are predicating colorhood of red. I do not think this is true. And neither did Frege (1950: sec. 57). It is certainly true that to say

(1) There are seventeen lions in the zoo

is not to predicate seventeen-hood of each individual lion. I suppose that it is also true that if there are seventeen lions in the zoo and also seventeen tigers in the zoo, the classes of lions-in-the-zoo and tigers-in-the-zoo are in a class together, though we shall return to that. It does not follow from this that (1) predicates seventeen-hood of one of those classes. First of all, the grammatical evidence for this is scanty indeed. The best one can conjure up by way of an example of the occurrence of a number word in predicative position is a rather artificial one like

(2) The lions in the zoo are seventeen.

If we do not interpret this as a statement about the ages of the beasts, we see that such statements do not predicate anything of any individual lion. One might then succumb to the temptation of analyzing (2) as the noun phrase "The lions in the zoo" followed by the verb phrase "are seventeen," where the analysis is parallel to that of

(3) The Cherokees are vanishing

where the noun phrase refers to the class and the verb phrase predicates something of that class. But the parallel is short-lived. For we soon notice that (2) probably comes into the language by deletion from

(4) The lions in the zoo are seventeen in number,

which in turn probably derives from something like

(5) Seventeen lions are in the zoo.

This is no place to explore in detail the grammar of number words. Suffice it to point out that they differ in many important respects from words we do not hesitate to call predicates. Probably the closest thing to a genuine class predicate involving number words is something on the model of "seventeen-membered" or "has seventeen members." But the step from there to "seventeen" being itself a predicate of classes is a long one indeed. In fact, I should think that pointing to the above two predicates gives away the show – for what is to be the analysis of "seventeen" as it occurs in those phrases?

Not only is there scanty grammatical evidence for this view, there

seems to be considerable evidence against it, as any scrutiny of the similarity of function among the number words and "many," "few," "all," "some," "any," and so forth will immediately reveal. The proper study of these matters will have to await another context, but the nonpredicative nature of number words can be further seen by noting how different they are from, say, ordinary adjectives, which do function as predicates. We have already seen that there are really no occurrences of number words in typical predicative position (that is, in "is (are) ..."), the only putative cases being along the lines of (2) above, and therefore rather implausible. The other anomaly is that number words normally outrank *all* adjectives (or all other adjectives, if one wants to class them as such) in having to appear at the head of an adjective string, and not inside. This is such a strong ranking that deviation virtually inevitably results in ungrammaticalness:

(6) The five lovely little square blue tiles

is fine, but any modification of the position of "five" yields an ungrammatical string; the farther to the right, the worse.[6]

Further reason for denying the predicative nature of number words comes from the traditional first-order analysis of sentences such as (1), with which we started. For that is usually analyzed as:

(7) $$(\exists x_1)\ldots(\exists x_{17})(Lx_1 \cdot Lx_2 \cdot \ldots \cdot Lx_{17} \cdot x_1 \neq x_2 \cdot x_1 \neq x_3 \cdot \ldots \cdot x_{16} \neq x_{17} \cdot (y)(Ly \supset . \, y = x_1 \vee y = x_2 \vee \ldots \vee y = x_{17})).$$

The only predicate in (1) which remains is "lion in the zoo," "seventeen" giving way to numerous quantifiers, truth functions, variables, and occurrences of "=," unless, of course, one wishes to consider these also to be predicates of classes. But there are slim grounds indeed for the view that (1) or (7) predicates seventeen-hood of the class of lions in the zoo. Number words function so much like operators such as "all," "some," and so forth, that a readiness to make class names of them should be accompanied by a readiness to make the corresponding move with respect to quantifiers, thereby proving (in traditional philosophic

[6] It might be thought that constructions such as

(i) The hungry five went home

constitute counterexamples to the thesis that number words must come first in an adjective string. But they do not. For in (i) and similar cases, the number word occurs as a noun, and not as an adjective, probably deriving from

(ii) The five hungry NP_{pl} went home

by the obvious transformation, and should be understood as such. There are certain genuine counterexamples, but the matter is too complicated for discussion here.

fashion) the existence of the one, the many, the few, the all, the some, the any, the every, the several, and the each.[7]

But then, what support *does* this view have? Well, this much: if two classes each have seventeen members, there probably exists a class which contains them both in virtue of that fact. I say "probably" because this varies from set theory to set theory. For example, this is not the case with type theory, since the two classes have both to be of the same type. But in no consistent theory is there a class of all classes with seventeen members, at least not alongside the other standard set-theoretical apparatus. The existence of the paradoxes is itself a good reason to deny to "seventeen" this univocal role of designating the class of all classes with seventeen members.

I think, therefore, that we may conclude that "seventeen" *need* not be considered a predicate of classes, and there is similarly no necessity to view 3 as the set of all triplets. This is not to deny that "is a class having three members" is a predicate of classes; but that is a different matter indeed. For that follows from all of the accounts under consideration.[8] Our present problem is to see if there is one account which can be established to the exclusion of all others, thereby settling the issue of which sets the numbers really are. And it should be clear by now that there is not. Any purpose we may have in giving an account of the notion of number and of the individual numbers, other than the question-begging one of proving of the right set of sets that *it* is the set of numbers, will be equally well (or badly) served by any one of the infinitely many accounts satisfying the conditions we set out so tediously. There is little need to examine all the possibilities in detail, once the traditionally favored one of Frege and Russell has been seen not to be uniquely suitable.

Where does that leave us? I have argued that at most one of the infinitely many different accounts satisfying our conditions can be correct, on the grounds that they are not even extensionally equivalent, and therefore at least all but one, and possibly all, contain conditions that are not necessary and that lead to the identification of the numbers with some particular set of sets. If numbers are sets, then they must be *particular sets,* for each set is some particular set. But if the number 3 is really one set rather than another, it must be possible to give some cogent reason for thinking so; for the position that this is an unknowable truth is hardly tenable. But there seems to be little to choose among the accounts. Relative to our purposes in giving an account of these matters, one will do as

[7]And indeed why not "I am the one who gave his all in fighting for the few against the many"?

[8]Within the bounds imposed by consistency.

well as another, stylistic preferences aside. There is no way connected with the reference of number words that will allow us to choose among them, *for the accounts differ at places where there is no connection whatever between features of the accounts and our uses of the words in question.* If all the above is cogent, then there is little to conclude except that any feature of an account that identifies 3 with a set is a superfluous one – and that therefore 3, and its fellow numbers, could not be sets at all.

III. Way out

In this third and final section, I shall examine and urge some considerations that I hope will lend plausibility to the conclusion of the previous section, if only by contrast. The issues involved are evidently so numerous and complex, and cover such a broad spectrum of philosophic problems, that in this paper I can do no more than indicate what I think they are and how, in general, I think they may be resolved. I hope nevertheless that a more positive account will emerge from these considerations.

A. Identity. Throughout the first two sections, I have treated expressions of the form

(8) $$n=s,$$

where n is a number expression and s a set expression as if I thought that they made perfectly good sense, and that it was our job to sort the true from the false.[9] And it might appear that I had concluded that all such statements were false. I did this to dramatize the kind of answer that a Fregean might give to the request for an analysis of number – to point up the kind of question Frege took it to be. For he clearly wanted the analysis to determine a truth value for each such identity. In fact, he wanted to determine a sense for the result of replacing s with any name or description whatsoever (while an expression ordinarily believed to name a number occupied the position of n). Given the symmetry and transitivity of identity, there were three kinds of identities satisfying these conditions, corresponding to the three kinds of expressions that can appear on the right:

(a) with some arithmetical expression on the right as well as on the left (for example, "$2^{17}=4{,}892$," and so forth);

[9]I was pleased to find that several of the points in my discussion of Frege have been made quite independently by Charles Parsons (1965). I am indebted to his discussion for a number of improvements.

(b) with an expression designating a number, but not in a standard arithmetical way, as "the number of apples in the pot," or "the number of F's" (for example, 7 = the number of the dwarfs);
(c) with a referring expression on the right which is of neither of the above sorts, such as "Julius Caesar," "[[∅]]" (for example, 17 = [[[∅]]]).

The requirement that the usual laws of arithmetic follow from the account takes care of all identities of the first sort. Adding an explication of the concept of cardinality will then suffice for those of kind (b). But to include those of kind (c), Frege felt it necessary to find some "objects" for number words to name and with which numbers could be identical. It was at this point that questions about which set of objects the numbers *really* were began to appear to need answering for, evidently, the simple answer "numbers" would not do. To speak from Frege's standpoint, there is a world of objects – that is, the designata or referents of names, descriptions, and so forth – in which the identity relation had free reign. It made sense for Frege to ask of *any* two names (or descriptions) whether they named the same object or different ones. Hence the complaint at one point in his argument that, thus far, one could not tell from his definitions whether Julius Caesar was a number.

I rather doubt that in order to explicate the use and meaning of number words one will have to decide whether Julius Caesar was (is?) or was not the number 43. Frege's insistence that this needed to be done stemmed, I think, from his (demonstrably) inconsistent logic (interpreted sufficiently broadly to encompass set theory). All items (names) in the universe were on a par, and the question whether two names had the same referent always presumably had an answer – yes or no. The inconsistency of the logic from which this stems is of course *some* reason to regard the view with suspicion. But it is hardly a refutation, since one might grant the meaningfulness of all identity statements, the existence of a universal set as the range of the relation, and still have principles of set existence sufficiently restrictive to avoid inconsistency. But such a view, divorced from the naïve set theory from which it stems, loses much of its appeal. I suggest, tentatively, that we look at the matter differently.

I propose to deny that all identities are meaningful, in particular to discard all questions of the form of (c) above as senseless or "unsemantical" (they are not totally senseless, for we grasp enough of their sense to explain why they are senseless). Identity statements make sense only in contexts where there exist possible individuating conditions. If an expression of the form "$x=y$" is to have a sense, it can be only in contexts where it is clear that both x and y are of some kind or category C, and

that it is the conditions which individuate things *as the same C* which are operative and determine its truth value. An example might help clarify the point. If we know x and y to be lampposts (possibly the same, but nothing in the way they are designated decides the issue) we can ask if they are *the same lamppost*. It will be their color, history, mass, position, and so forth which will determine if they are indeed the same lamppost. Similarly, if we know z and w to be numbers, then we can ask if they are *the same number*. And it will be whether they are prime, greater than 17, and so forth which will decide if they are indeed the same number. But just as we cannot individuate a lamppost in terms of these latter predicates, neither can we individuate a number in terms of its mass, color, or similar considerations. What determines that something is a *particular lamppost* could not individuate it as a *particular number*. I am arguing that questions of the identity of a particular "entity" do not make sense. "Entity" is too broad. For such questions to make sense, there must be a well-entrenched predicate C, in terms of which one then asks about the identity of a *particular C*, and the conditions associated with identifying C's as *the same C* will be the deciding ones. Therefore, if for two predicates F and G there is no third predicate C which subsumes both and which has associated with it some uniform conditions for identifying two putative elements as the same (or different) C's, the identity statements crossing the F and G boundary will not make sense.[10] For example, it will make sense to ask of something x (which is in fact a chair) if it is the same . . . as y (which in fact is a table). For we can fill the blank with a predicate, "piece of furniture," and we know what it is for a and b to be the same or different pieces of furniture. To put the point differently, questions of identity contain the presupposition that the "entities" inquired about both belong to some general category. This presupposition is normally carried by the context or theory (that is, a more systematic context). To say that they are both "entities" is to make no presuppositions at all – for everything purports to be at least that. "Entity," "thing," "object" are words having a role in the language; they are place fillers whose function is analogous to that of pronouns (and, in more formalized contexts, to variables of quantification).

Identity *is* id-entity, but only within narrowly restricted contexts. Alternatively, what constitutes an entity is category or theory dependent. There are really two correlative ways of looking at the problem. One

[10] To give a precise account, it will be necessary to explain "uniform conditions" in such a way as to rule out the obvious counterexamples generated by constructed *ad hoc* disjunctive conditions. But to discuss the way to do this would take us too far afield. I do not pretend to know the answer in any detail.

might conclude that identity is systematically ambiguous, or else one might agree with Frege, that identity is unambiguous, always meaning sameness of object, but that (contra-Frege now) the notion of an *object* varies from theory to theory, category to category – and therefore that his mistake lay in failure to realize this fact. This last is what I am urging, for it has the virtue of preserving identity as a general logical relation whose application in any given well-defined context (that is, one within which the notion of object is univocal) remains unproblematic. Logic can then still be seen as the most general of disciplines, applicable in the same way to and within any given theory. It remains the tool applicable to all disciplines and theories, the difference being only that it is left to the discipline or theory to determine what shall count as an "object" or "individual."

That this is not an implausible view is also suggested by the language. Contexts of the form "the same *G*" abound, and indeed it is in terms of them that identity should be explained, for what will be counted as the same *G* will depend heavily on *G*. The same *man* will have to be an individual man; "the same *act*" is a description that can be satisfied by many individual acts, or by only one, for the individuating conditions for acts make them sometimes types, sometimes tokens. Very rare in the language are contexts open to (satisfiable by) any kind of "thing" whatsoever. There are some – for example, "Sam referred to . . . ," "Helen thought of . . ." – and it seems perfectly all right to ask if what Sam referred to on some occasion was what Helen thought of. But these contexts are very few, and they all seem to be intensional, which casts a referentially opaque shadow over the role that identity plays in them.

Some will want to argue that identities of type (c) are not senseless or unsemantical, but simply false – on the grounds that the distinction of categories is one that cannot be drawn. I have only the following argument to counter such a view. It will be just as hard to explain how one *knows* that they are false as it would be to explain how one knows that they are senseless, for normally we know the falsity of some identity "$x=y$" only if we know of x (or y) that it has some characteristic that we know y (or x) *not* to have. I know that $2 \neq 3$ because I know, for example, that 3 is odd and 2 is not, yet it seems clearly wrong to argue that we know that $3 \neq [[[\emptyset]]]$ because, say, we know that 3 has no (or seventeen, or infinitely many) members while $[[[\emptyset]]]$ has exactly one. We know no such thing. We do not know that it does. But that does not constitute knowing that it does not. What is enticing about the view that these are all false is, of course, that they hardly seem to be open questions to which we may find the answer any day. Clearly, all the evidence is in; if no decision is possible on the basis of it, none will ever be possible. But for the purposes at hand the difference between these two views is not a very serious one. I should certainly be happy with the conclusion that all identities of type (c) are either senseless or false.

B. Explication and reduction. I would like now to approach the question from a slightly different angle. Throughout this paper, I have been discussing what was substantially Frege's view, in an effort to cast some light on the meaning of number words by exposing the difficulties involved in trying to determine which objects the numbers really are. The analyses we have considered all contain the condition that numbers are sets, and that therefore each individual number is some individual set. We concluded at the end of Section II that numbers could not be sets at all – on the grounds that there are no good reasons to say that any particular number is some particular set. To bolster our argument, it might be instructive to look briefly at two activities closely related to that of stating that numbers *are* sets – those of explication and reduction.

In putting forth an explication of number, a philosopher may have as part of his explication the statement that $3=[[[\emptyset]]]$. Does it follow that he is making the kind of mistake of which I accused Frege? I think not. For there is a difference between *asserting* that 3 *is* the set of all triplets and *identifying* 3 with that set, which last is what might be done in the context of some explication. I certainly do not wish what I am arguing in this paper to militate against identifying 3 with anything you like. The difference lies in that, normally, one who identifies 3 with some particular set does so for the purpose of presenting some theory and does not claim that he has *discovered* which object 3 really is. We might want to know whether some set (and relations and so forth) would do as number surrogates. In investigating this it would be entirely legitimate to state that making such an identification we can do with that set (and those relations) what we now do with the numbers. Hence we find Quine saying:

> Frege dealt with the question "What is a number?" by showing how the work for which the objects in question might be wanted could be done by objects whose nature was presumed to be less in question. (Quine 1960: 262)

Ignoring whether this is a correct interpretation of Frege, surely someone who says this would not claim that, since the answer turned out to be "Yes," it is now clear that numbers had really been sets all along. In such a context, the adequacy of some system of objects to the task is a very real question and one which can be settled. Under our analysis, *any* system of objects, sets or not, that forms a recursive progression must be adequate. Thus, discovering that a different system will do the job very nicely cannot be to discover which objects the numbers are .... Explication in the above reductionistic sense, is therefore neutral with respect to the sort of problem we have been discussing, but it does cast some sobering light on what it is to be an individual number.

There is another reason to deny that it would be legitimate to use the reducibility of arithmetic to set theory as a reason to assert that numbers are really sets after all. Gaisi Takeuti has shown that the Gödel-von Neumann-Bernays set theory is in a strong sense *reducible to* the theory of ordinal numbers less than the least inaccessible number (1954). No wonder numbers are sets; sets are really (ordinal) numbers, after all. *But now, which is really which?*

These brief comments on reduction, explication, and what they might be said to achieve in mathematics lead us back to the quotation from Richard Martin which heads this paper. Martin correctly points out that the mathematician's interest stops at the level of structure. If one theory can be modeled in another (that is, reduced to another) then further questions about whether the individuals of one theory are really those of the second just do not arise. In the same passage, Martin goes on to point out (approvingly, I take it) that the philosopher is not satisfied with this limited view of things. He wants to know more and does ask the questions in which the mathematician professes no interest. I agree. He does. And mistakenly so. It will be the burden of the rest of this paper to argue that such questions miss the point of what arithmetic, at least, is all about.

C. Conclusion: numbers and objects. It was pointed out above that any system of objects, whether sets or not, that forms a recursive progression must be adequate. But this is odd, for any recursive set can be arranged in a recursive progression. So what matters, really, is not any condition on the *objects* (that is, on the set) but rather a condition on the relation under which they form a progression. To put the point differently – and this is the crux of the matter – that any recursive sequence whatever would do suggests that what is important is not the individuality of each element but the structure which they jointly exhibit. This is an extremely striking feature. One would be led to expect from this fact alone that the question of whether a particular "object" – for example, [[[Ø]]] – would do as a replacement for the number 3 would be pointless in the extreme, as indeed it is. "Objects" do not do the job of numbers singly; the whole system performs the job or nothing does. I therefore argue, extending the argument that led to the conclusion that numbers could not be sets, that numbers could not be objects at all; for there is no more reason to identify any individual number with any one particular object than with any other (not already known to be a number).

The pointlessness of trying to determine which objects the numbers are thus derives directly from the pointlessness of asking the question of any individual number. For arithmetical purposes the properties of numbers which do not stem from the relations they bear to one another in virtue of being arranged in a progression are of no consequence whatsoever. But it would be only these properties that would single out a number as this object or that.

Therefore, numbers are not objects at all, because in giving the properties (that is, necessary and sufficient) of numbers you merely characterize an *abstract structure* – and the distinction lies in the fact that the "elements" of the structure have no properties other than those relating them to other "elements" of the same structure. If we identify an abstract structure with a system of relations (in intension, of course, or else with the set of all relations in extension isomorphic to a given system of relations), we get arithmetic elaborating the properties of the "less-than" relation, or of all systems of objects (that is, *concrete* structures) exhibiting that abstract structure. That a system of objects exhibits the structure of the integers implies that the elements of that system have some properties not dependent on structure. It must be possible to individuate those objects independently of the role they play in that structure. But this is precisely what cannot be done with the numbers. To *be* the number 3 is no more and no less than to be preceded by 2, 1, and possibly 0, and to be followed by 4, 5, and so forth. And to *be* the number 4 is no more and no less than to be preceded by 3, 2, 1, and possibly 0, and to be followed by .... *Any* object can *play the role of* 3; that is, any object can be the third element in some progression. What is peculiar to 3 is that it defines that role – not by being a paradigm of any object which plays it, but by representing the relation that any third member of a progression bears to the rest of the progression.

Arithmetic is therefore the science that elaborates the abstract structure that all progressions have in common merely in virtue of being progressions. It is not a science concerned with particular objects – the numbers. The search for which independently identifiable particular objects the numbers really are (sets? Julius Caesars?) is a misguided one.

On this view many things that puzzled us in this paper seem to fall into place. Why so many interpretations of number theory are possible without any being uniquely singled out becomes obvious: there is no unique set of objects that are the numbers. Number theory is the elaboration of the properties of *all* structures of the order type of the numbers. The number words do not have single referents. Furthermore, the reason identification of numbers with objects works wholesale but fails utterly object by object is the fact that the theory is elaborating an abstract structure and not the properties of independent individuals, any one of which could be characterized without reference to its relations to the rest. Only when we are considering a particular sequence as being, not the numbers, but *of the structure of the numbers* does the question of which element is, or rather *corresponds to*, 3 begin to make any sense.

Slogans like "Arithmetic is about numbers," "Number words refer to numbers," when properly urged, may be interpreted as pointing out two quite distinct things: (1) that number words are not names of special non-numerical entities, like sets, tomatoes, or Gila monsters; and (2) that a purely formalistic view that fails to assign any meaning whatsoever to the statements of number theory is also wrong. They need not be incompatible with what I am urging here.

This last formalism is too extreme. But there is a modified form of it, also denying that number words are names, which constitutes a plausible and tempting extension of the view I have been arguing. Let me suggest it here. On this view the sequence of number words is just that – a sequence of words or expressions with certain properties. There are not two kinds of things, numbers and number words, but just one, the words themselves. Most languages contain such a sequence, and any such sequence (of words or terms) will serve the purposes for which we have ours, provided it is recursive in the relevant respect. In counting, we do not correlate sets with initial segments of the numbers as extralinguistic entities, but correlate sets with initial segments of the sequence of number *words*. The central idea is that this recursive sequence is a sort of yardstick which we use to measure sets. Questions of the identification of the referents of number words should be dismissed as misguided in just the way that a question about the referents of the parts of a ruler would be seen as misguided. Although any sequence of expressions with the proper structure would do the job for which we employ our present number words, there is still some reason for having one, relatively uniform, notation: ordinary communication. Too many sequences in common use would make it necessary for us to learn too many different equivalences. The usual objection to such an account – that there is a distinction between numbers and number words which it fails to make will, I think, not do. It is made on the grounds that "two," "*zwei*," "*deux*," "2" are all supposed to "stand for" the same number but yet are *different* words (one of them not a word at all). One can mark the differences among the expressions in question, and the similarities as well, without conjuring up some extralinguistic objects for them to name. One need only point to the similarity of function: within any numbering system, what will be

important will be what place in the system any particular expression is used to mark. All the above expressions share this feature with one another – and with the binary use of "10," but not with its decimal employment. The "ambiguity" of "10" is thus easily explained. Here again we see the series-related character of individual numbers, except that it is now mapped a little closer to home. One cannot tell what number a particular expression represents without being given the sequence of which it forms a part. It will then be from its place in that sequence – that is, from its relation to other members of the sequence, *and from the rules governing the use of the sequence in counting* – that it will derive its individuality. It is for this last reason that I urged, contra Quine, that the account of cardinality must explicitly be included in the account of number (see note 3).

Furthermore, other things fall into place as well. The requirement, discussed in Section I, that the "less-than" relation be recursive is most easily explained in terms of a recursive notation. After all, the whole theory of recursive functions makes most sense when viewed in close connection with notations rather than with extralinguistic objects. This makes itself most obvious in three places: the development of the theory by Post systems, by Turing machines, and in the theory of constructive ordinals, where the concern is frankly with recursive notations for ordinals. I do not see why this should not be true of the finite ordinals as well. For a set of *numbers* is recursive if and only if a machine of a particular sort could be programmed to generate them in order of magnitude – that is, to generate the standard or canonical notations for those numbers following the (reverse) order of the "less-than" relation. If that relation over the notation were not recursive, the above theorem would not hold.

It also becomes obvious why every analysis of number ever presented has had a recursive "less-than" relation. If what we are generating is a notation, the most natural way for generating it is by giving recursive rules for getting the next element from any element you may have – and you would have to go far out of your way (and be slightly mad) to generate the notation and then define "less than" as I did on page 276, above, in discussing the requirement of recursiveness.

Furthermore, on this view, we learn the elementary arithmetical operations as the cardinal operations on small sets, and extend them by the usual algorithms. Arithmetic then becomes cardinal arithmetic at the earlier levels in the obvious way, and the more advanced statements become easily interpretable as *projections* via truth functions, quantifiers, and the recursive rules governing the operations. One can therefore be this sort of formalist without denying that there is such a thing as arithmetical truth other than derivability within some given system. One can even explain what the ordinary formalist apparently cannot – why these axioms were chosen and which of two possible consistent extensions we should adopt in any given case.

But I must stop here. I cannot defend this view in detail without writing a book. To return in closing to our poor abandoned children, I think we must conclude that their education was badly mismanaged – not from the mathematical point of view, since we have concluded that there is no mathematically significant difference between what they were taught and what ordinary mortals know, but from the philosophical point of view. They think that numbers are really sets of sets while, if the truth be known, there are no such things as numbers; which is not to say that there are not at least two prime numbers between 15 and 20.

Mathematics without foundations

HILARY PUTNAM

Philosophers and logicians have been so busy trying to provide mathematics with a 'foundation' in the past half-century that only rarely have a few timid voices dared to voice the suggestion that it does not need one. I wish here to urge with some seriousness the view of the timid voices. I don't think mathematics is unclear; I don't think mathematics has a crisis in its foundations; indeed, I do not believe mathematics either has or needs 'foundations'. The much touted problems in the philosophy of mathematics seem to me, without exception, to be problems internal to the thought of various system builders. The systems are doubtless interesting as intellectual exercises; debate between the systems and research within the systems doubtless will and should continue; but I would like to convince you (of course I won't, but one can always hope) that the various systems of mathematical philosophy, without exception, need not be taken seriously.

By way of comparison, it may be salutory to consider the various 'crises' that philosophy has pretended to discover in the past. It is impressive to remember that at the turn of the century there was a large measure of agreement among philosophers – far more than there is now – on certain fundamentals. Virtually all philosophers were idealists of one sort or another. But even the nonidealists were in a large measure of agreement with the idealists. It was generally agreed any property of material objects – say, *redness* or *length* – could be ascribed to the object, if at all, only as a power to produce certain sorts of sensory experiences. When the man on the street thinks of a material object, according to this traditional view, he really thinks of a subjective object, not a real 'external' object. If there are external objects, we cannot really imagine what they are like; we know and can conceive only their powers. Either there are no external objects at all (Berkeley) – i.e. no objects 'external' to minds and their ideas – or there are, but they are *Dinge an sich*. In sum, then, philosophy flattered itself to have discovered not just a crisis, but a fundamental mistake, not in some special science, but in our most common-sense convictions about material objects. To put it

Reprinted with the kind permission of the editors from the *Journal of Philosophy* 64 (1967): 5–22.

crudely, philosophy thought itself to have shown that no one has ever really perceived a material object and that, if material objects exist at all (which was thought to be highly problematical), then no one *could* perceive, or even imagine, one.

Anyone maintaining at the turn of the century that the notions 'red' and 'hard' (or, more abstractly 'material object') were reasonably clear notions; that redness and hardness are *non*dispositional properties of material objects; that we see red things and see *that* they are red; and that *of course* we can imagine red objects, know what a red object is, etc., would have seemed unutterably foolish. After all, the most brilliant philosophers in the world all found difficulties with these notions. Clearly, the man is just too stupid to see the difficulties. Yet today this 'stupid' view is the view of many sophisticated philosophers, and the increasingly prevalent opinion is that it was the arguments purporting to show a contradiction in the view, and not the view itself, that were profoundly wrong. Moral: not everything that passes – in philosophy anyway – as a difficulty with a concept is one. And second moral: the fact that philosophers all agree that a notion is 'unclear' doesn't mean that it *is* unclear.

More recently there was a large measure of agreement among philosophers of science – far more than there is now – that, in some sense, talk about theoretical entities and physical magnitudes is 'highly derived talk' which, in the last analysis, reduces to talk about observables. Just a few years ago, we were being told that 'electron' is a 'partially interpreted' term, whereas 'red' is 'completely interpreted'. Today it is becoming increasingly clear that 'electron' is a term that has complete 'meaning' in every sense in which 'red' has 'meaning'; that the 'purpose' of talk about electrons is not simply to make successful predictions in observation language any more than the 'purpose' of talk about red things is to make true deductions about electrons; and that the whole question about how we 'introduce' theoretical terms was a mare's nest. I refrain from drawing another moral.

Today there is a large measure of agreement among philosophers of mathematics that the concept of a 'set' is unclear. I hope the above short review of some history of philosophy will indicate why I am less than overawed by this agreement. When philosophy discovers something wrong with science, sometimes science has to be changed – Russell's paradox comes to mind, as does Berkeley's attack on the actual infinitesimal – but more often it is philosophy that has to be changed. I do not think that the difficulties that philosophy finds with classical mathematics today are genuine difficulties; and I think that the philosophical interpretations of mathematics that we are being offered on every hand

are wrong, and that 'philosophical interpretation' is just what mathematics doesn't need. And I include my own past efforts in this direction.

I do not, however, mean to disparage the value of philosophical inquiry. If philosophy got itself into difficulties with the concept of a material object, it also got itself out; and the result is some modest but significant increase in our clarity about perception and knowledge. It is this sort of clarity about mathematical truth, mathematical 'objects', and mathematical necessity that I should like to see us attain; but I do not think the famous 'isms' in the philosophy of mathematics represent the road to that clarity. Let us therefore make a fresh start.

A sketch of my view

I think that the least mystifying way for me to discuss this topic is as follows: first to give a very cursory and superficial sketch of my own views, so that you will at least be able to guess at the positive position that underlies my criticism of others, and then to survey the alleged difficulties in set theory. Of course, any philosopher hates ever to say briefly, let alone superficially, what his own view on any topic is (although he is delighted to give such a statement to the view of any philosopher with whom he disagrees), because a superficial statement may make his view seem naive or even downright stupid. But such a statement is a great help to others, at least in getting an initial orientation, and for that reason I shall accept the risk involved.

In my view the chief characteristic of mathematical propositions is the very wide variety of equivalent formulations that they possess. I don't mean this in the trivial sense of cardinality: of course, every proposition possesses infinitely many equivalent formulations; what I mean is rather that in mathematics the number of ways of expressing what is in some sense the same fact (if the proposition is true) while apparently not talking about the same objects is especially striking.

The same situation does sometimes arise in empirical science, that is, the situation that what is in some sense the same fact can be expressed in two strikingly different ways, the most famous example being wave-particle duality in quantum mechanics. Reichenbach coined the happy expression 'equivalent descriptions' for this situation. The description of the world as a system of particles, not in the classical sense but in the peculiar quantum-mechanical sense, may be associated with a different picture than the description of the world as a system of waves, again not in the classical sense but in the quantum-mechanical sense; but the two theories are thoroughly intertranslatable, and should be viewed as having the same physical content. The same fact can be expressed either by saying that the electron is a wave with a definite wavelength λ or by saying that the electron is a particle with a sharp momentum p and an indeterminate position. What 'same fact' comes to here is, I admit, obscure. Obviously what is *not* being claimed is *synonymy* of *sentences*. It would be absurd to claim that the *sentence* 'there is an electron-wave with the wavelength λ' is *synonymous* with the *sentence* 'there is a particle electron with the momentum h/λ and a totally indeterminate position'. What is rather being claimed is this: that the two theories are compatible, not incompatible, given the way in which the theoretical primitives of each theory are now being understood; that indeed, they are not merely compatible but equivalent: the primitive terms of each admit of definition by means of the primitive terms of the other theory, and then each theory is a deductive consequence of the other. Moreover, there is no particular advantage to taking one of the two theories as fundamental and regarding the other one as *derived*. The two theories are, so to speak, on the same explanatory level. Any fact that can be explained by means of one can equally well be explained by means of the other. And in view of the systematic equivalence of statements in the one theory with statements in the other theory, there is no longer any point to regarding the formulation of a given fact in terms of the notions of one theory as more fundamental than (or even as *significantly* different from) the formulation of the fact in terms of the notions of the other theory. In short, what has happened is that the systematic equivalences between the sentences of the two theories have become so well known that they *function* virtually as synonymies in the actual practice of science.

Of course, the fact that two theories can be related in this way is not by itself either surprising or important. It would not be worth remarking that two theories are related in this way if the pictures associated with the two theories were not apparently incompatible or at least very different. In mathematics, the different equivalent formulations of a given mathematical proposition do not call to mind apparently *incompatible* pictures as do the different equivalent formulations of the quantum theory, but they do sometimes call to mind radically different pictures, and I think that the way in which a given philosopher of mathematics proceeds is often determined by which of these pictures he has in mind, and this in turn is often determined by which of the equivalent formulations of the mathematical propositions with which he deals he takes as primary.

Of the many possible 'equivalent descriptions' of the realm of mathematical facts, there are two which seem to me to have especial importance. I shall refer to these, somewhat misleadingly, I admit, by the titles 'Mathematics as Modal Logic' and 'Mathematics as Set Theory', The

second, I take it, needs no explanation. Everyone is today familiar with the conception of mathematics as the description of a 'universe' of 'mathematical objects' – and, in particular, with the conception of mathematics as describing relations among *sets*. However, the picture would not be significantly different if one said 'sets and numbers' – that numbers can themselves be 'identified' with sets seems today a matter of minor importance; the important thing about the picture is that mathematics describes 'objects'. The other conception is less familiar, and I shall say a few words about it.

Consider the assertion that there is a counterexample to Fermat's 'last theorem'; i.e. that there is an nth power which is the sum of two nth powers, $2<n$, all three numbers positive. Abbreviate the standard formula that expresses this statement in first-order arithmetic as '$\sim$*Fermat*'. If $\sim$*Fermat* is provable, then, in fact, $\sim$*Fermat* is provable already from a certain easily specified finite subset of the theorems of first-order arithmetic. (N.B., this is owing to the fact that it takes only one counterexample to refute a generalization. So the portion of first-order arithmetic in which we can prove all true statements of the form $x^n+y^n\neq z^n$, x, y, z, n *constant* integers, is certainly strong enough to *disprove* Fermat's last theorem if the last theorem be false, notwithstanding the fact that *all* of first-order arithmetic may be too weak to *prove* Fermat's last theorem if the last theorem be true. And the portion of first-order arithmetic just alluded to is known to be finitely axiomatizable.) Let 'AX' abbreviate the conjunction of the axioms of the finitely axiomatizable subtheory of first-order arithmetic just alluded to. Then Fermat's last theorem is *false* just in case '$AX \supset \sim Fermat$' is valid, i.e. just in case

(1) $$\Box(AX \supset \sim Fermat)$$

Since the truth of (1), in case (1) *is* true, does not depend upon the meaning of the arithmetical primitives, let us suppose these to be replaced by 'dummy letters' (predicate letters). To fix our ideas imagine that the primitives in terms of which AX and $\sim$*Fermat* are written are the two three-term relations 'x is the sum of y and z' and 'x is the product of y and z' (exponentiation is known to be first-order-definable from these, and so, of course, are *zero* and *successor*). Let $\text{AX}(S, T)$ and $\sim\text{FERMAT}(S, T)$ be like AX and $\sim$*Fermat* except for containing the 'dummy' triadic predicate letters S, T, where AX and $\sim$*Fermat* contain the constant predicates 'x is the sum of y and z' and 'x is the product of y and z'. Then (1) is essentially a truth of pure modal logic (if it is true), since the constant predicates occur 'inessentially'; and this can be brought out by replacing (1) by the abstract schema:

(2) $$\Box[\text{AX}(S, T) \supset \sim \text{FERMAT}(S, T)]$$

– and this is a schema of pure first-order modal logic.

Now then, the mathematical content of the assertion (2) is certainly the same as that of the assertion that *there exist numbers* x, y, z, n $(2<n, x, y, z\neq 0)$ such that $x^n+y^n=z^n$. Even if the expressions involved are not synonymous, the mathematical equivalence is so obvious that they might as well be synonymous, as far as the mathematician is concerned. Yet the pictures in the mind called up by these two ways of formulating what one might as well consider to be the same mathematical assertion can be quite different. When one speaks of the 'existence of numbers' one gets the picture of mathematics as describing eternal objects; while (2) simply says that $\text{AX}(S, T)$ entails $\text{FERMAT}(S, T)$, no matter how one may interpret the predicate letters 'S' and 'T', and this scarcely seems to be about 'objects' at all. Of course, one can strain after objects if one wants. One can, for example, interpret the dummy letters 'S' and 'T' as quantifiers over 'the totality of all properties', if one wishes. But this is hardly necessary, since one can find a particular substitution instance of (2), even in a nominalistic language (apart from the '$\Box$') which is equivalent to (2) (just choose predicates S^* and T^* to put for S and T such that it is not mathematically impossible that the objects in their field should form an ω-sequence, and such that, if the objects in their field did form an ω-sequence, S^* would be isomorphic to addition of integers, and T^* to multiplication, in the obvious sense). Or one can interpret '$\Box$' as a predicate of statements, rather than as a statement connective, in which case what (2) asserts is that a certain object, namely the statement '$\text{AX}(S, T) \supset \sim \text{FERMAT}(S, T)$' has a certain property ('being necessary'). But still, the only 'object' this commits us to is the statement '$\text{AX}(S, T) \supset \sim \text{FERMAT}(S, T)$', and one has to be pretty compulsive about one's nominalistic cleanliness to scruple about *this*. In short, if one fastens on the first picture (the 'object' picture), then mathematics is wholly extensional, but presupposes a vast totality of eternal objects; while if one fastens on the second picture (the 'modal' picture), then mathematics has *no* special objects of its own, but simply tells us what follows from what. If 'Platonism' has appeared to be *the* issue in the philosophy of mathematics of recent years, I suggest that it is because we have been too much in the grip of the first picture.

So far I have only indicated how one very special mathematical proposition can be treated as a statement involving modalities, but not special objects. I believe that, by making a more complex and iterated use of modal notions, one can analyze the notion of *a standard model for set theory*, and thus extend the objects – modalities duality that I am discussing to the whole of classical mathematics. I shall not show this now; but, needless to say, I would not deal at such length with this one special example if I did not believe it to represent, in some sense, the general situation. For the moment, I shall ask you to accept it on faith that this extension to the general case can be carried out.

What follows, I believe, is that each of these two ways of looking at mathematics can be used to clarify the other. If one is puzzled by the modalities (and I am concerned here with necessity in Quine's narrower sense of logical validity, excluding necessities that depend on alleged synonymy relations in natural languages), then one can be helped by the set-theoretic notion of a *model* (necessity = truth in all models; possibility = truth in some model). On the other hand, if one is puzzled by the question recently raised by Benacerraf (1965; reprinted in this volume): how numbers can be 'objects' if they have *no* properties except order in a particular ω-sequence, then, I believe, one can be helped by the answer: call them 'objects' if you like (they *are* objects in the sense of being things one can quantify over); but remember that these objects have the special property that each fact about them is, in an equivalent formulation, simply a fact about *any* ω-sequence. 'Numbers exist'; but all this comes to, for mathematics anyway, is that (1) ω-sequences are *possible* (mathematically speaking); and (2) there are *necessary* truths of the form 'if α is an ω-sequence, then...' (whether any *concrete* example of an ω-sequence exists or not). Similarly, there is not, from a mathematical point of view, any significant difference between the assertion that *there exists a set of integers* satisfying an arithmetical condition and the assertion that *it is possible to select* integers so as to satisfy the condition. Sets, if you will forgive me for parodying John Stuart Mill, are permanent possibilities of selection.

The question of decidability

The sense that there is a 'crisis in the foundations' of mathematics has many sources. Morris Kline cites the development of non-Euclidean geometry (which shook the idea that the axioms of a mathematical discipline must be *truths*), the lack of a consistency proof for mathematics, and the lack of a universally acceptable solution to the antinomies. In addition to these, one might mention Gödel's theorem (Kline does mention it, in fact, in connection with the consistency problem). For Gödel's theorem suggets that the truth or falsity of some mathematical statements might be impossible in principle to ascertain, and this has led some to wonder if we even know what we mean by 'truth' and 'falsity' in such a context.

Now, the example of non-Euclidean geometry does show, I believe, that our notions of what is 'self-evident' have to be subject to revision, not just in the light of new observations, but in the light of new *theories*. The intuitive evidence for the proposition that two lines cannot be a constant distance apart for half their length (i.e. in one half-plane) and then start to approach each other (as geodesics can in General Relativity, e.g. light rays which come in from infinity parallel and then approach each other as they pass on opposite sides of the sun) is as great as the intuitive evidence for the axioms of number theory. I believe that under certain circumstances revisions in the axioms of arithmetic, or even of propositional calculus (e.g. the adoption of a modular logic as a way out of the difficulties in quantum mechanics), is fully conceivable. The philosophical ploy which consists in saying 'then terms would have changed meaning' is uninteresting – except as a question in the philosophy of linguistics, of course – unless one can show that in their 'old meaning' the sentences of the theory in question can still (after the transition to non-Euclidean geometry, or non-Archimedean arithmetic, or modular logic) be admitted to have formerly expressed propositions that are clear and true. If in some sense there are 'Euclidean straight lines' in our space, then the transition to, say, Riemannian geometry *could* (not necessarily *should*) be regarded as a mere 'change of meaning'. But (1) there are *no* curves in space (if the world is Riemannian) that satisfy Euclid's theorems about straight lines; and (2) even if the world is Lobatchevskian, there are no *unique* such curves – to choose any particular remetricization which leads to Euclidean geometry and say '*this* is what "distance", "straight line", etc., *used* to mean' would be arbitrary. In short, the price one pays for the adoption of non-Euclidean geometry is to deny that there are *any* propositions which might *plausibly* have been in the minds of the people who believed in Euclidean geometry and which are simultaneously clear and true. Similarly, if one accepts the interpretation of quantum mechanics that is based on modular logic, then one has to deny that there has been a change in the meaning of the relevant sentences, or else deny that there are any unique propositions which might have been in the minds of those who formerly used those sentences and which were both clear and true. You can't have your conceptual revolution and minimize it too!

Yet all this does not, I think, mean that there is a crisis in the foundations of mathematics. It does not even mean that mathematics becomes an empirical science in the ordinary sense of that term. For the chief characteristic of empirical science is that for each theory there are usually alternatives in the field, or at least alternatives struggling to be born. As long as the major parts of classical logic and number theory and analysis

have no alternatives in the field – alternatives which require a change in the axioms and which effect the simplicity of total science, including empirical science, so that a choice has to be made – the situation will be what it has always been. We will be justified in accepting classical propositional calculus or Peano number theory not because the relevant statements are 'unrevisable in principle' but because a great deal of science presupposes these statements and because no real alternative is in the field. Mathematics, on this view, does become 'empirical' in the sense that one is allowed to try to *put* alternatives into the field. Mathematics can be wrong, and not just in the sense that the proofs might be fallacious or that the axioms might not (if we reflected more deeply) be really self-evident. Mathematics (or rather, some mathematical theory) might be wrong in the sense that the 'self-evident' axioms might be false, and the axioms that are true might not be 'evident' at all. But this does not make the pursuit of truth impossible in mathematics any more than it has in empirical science, nor does it mean that we should not trust our intuitions when we have nothing better to go on. After all, a mathematical theory that has become the basis of a successful and powerful scientific system, including many important empirical applications, is not being accepted *merely* because it is 'intuitive', and if someone objects to it we have the right to say 'propose something better!' What this does do, rather, is make the 'foundational' view of mathematical knowledge as suspect as the 'foundational' view of empirical knowledge (if one cares to retain the 'mathematical–empirical' distinction at all).

Again, I cannot weep bitter tears about the lack of a consistency proof for classical mathematics. Even if such a proof were possible, it would only be a development within mathematics and not a foundation for mathematics. Not only would it be possible to raise philosophical questions about the branch of mathematics that was used for the consistency proof; but, in any case, science demands much more of a mathematical theory than that it should merely be *consistent*, as the example of the various alternative systems of geometry already dramatizes.

The question of the significance of the antinomies, and of what to do about the existence of several different approaches to overcoming them, is far more difficult. I propose to defer this question for a moment and to consider first the significance of Gödel's theorem and, more generally, of the existence of mathematically undecidable propositions.

Strictly speaking, all Gödel's theorem shows is that, in any particular consistent axiomatizable extension of certain finitely axiomatizable subtheories of Peano arithmetic, there are propositions of number theory that can neither be proved nor disproved. (I think it is fair to call this 'Gödel's theorem', even though this statement of it incorporates strengthenings due to Rosser and Tarski, Mostowski, Robinson.) It does not follow that any proposition of number theory is, in some sense, absolutely undecidable. However, it may well be the case that some proposition of elementary number theory is neither provable nor refutable in any system whose axioms rational beings will ever have any good reason to accept. This has caused some to doubt whether every mathematical proposition, or even every proposition of the elementary theory of numbers, can be thought of as having a truth value.

A similar consideration is raised by Paul Cohen's recent work in set theory, when that work is taken together with Gödel's classical relative consistency proof of the axiom $V=L$ (which implies the axiom of choice and the generalized continuum hypothesis). Together these results of Gödel and Cohen establish the full independence of the continuum hypothesis (for example) from the other axioms of set theory, assuming those other axioms to be consistent. A striking feature of both proofs is their invariance under small (or even moderately large) perturbations of the axioms. It appears quite possible today that no decisive consideration will ever appear (such as a set-theoretic axiom we have 'overlooked') which will reveal that a system in which the continuum hypothesis is provable is the correct one, and that no consideration will ever appear which will reveal that a system in which the continuum hypothesis is refutable is the correct one. In short, the truth value of the continuum hypothesis – assuming it has a truth value – may be undiscoverable by rational beings, or at least by the 'rational beings' that actually do exist, or ever will exist. Then, what reason is there to think that it has a truth value?

This 'argument' is sometimes taken to show that the notion of a set is unclear. For, since the argument 'shows' (sic!) that the continuum hypothesis has no truth value and the continuum hypothesis involves the concept of a set, the only plausible explanation of the truth-value failure is some unclarity in the notion of a set. (It would be an interesting exercise to find *all* the faults in this particular bit of reasoning. It is horrible, isn't it?)

The first point to notice is that the existence of propositions whose truth value we have no way of discovering is not at all peculiar to mathematics. Consider the assertion that there are infinitely many binary stars (considering the entire space–time universe, i.e. counting binary stars past, present, and future). It is not at all clear that we can discover the truth value of this assertion. Sometimes it is argued that such an assertion is 'verifiable (or at least confirmable) in principle', because it may *follow from a theory*. It is true that in one case we can discover the truth

value of this proposition. Namely, if either it or its negation is derivable from laws of nature that we can confirm, then its truth value can be discovered. But it could just happen that there are infinitely many binary stars, without this being required by any law. Moreover, the distribution might be quite irregular, so that ordinary statistical inference could not discover it. Indeed, at some point I cease to understand the question 'Is it always possible *in principle* to discover the truth value of this proposition?' – for the methods of inquiry permitted ('inductive' methods) are just too ill defined a set. But I suspect that, given any *formalizable* inductive logic, one could describe a logically possible world in which (1) there were infinitely many binary stars; and (2) one could never discover this fact using that inductive logic. (Of course, the argument that the proposition is 'confirmable in principle' because it could follow from a theory does not even purport to show that in every possible world the truth or falsity of this statement could be induced from a finite amount of observational material using some inductive method; rather it shows that in *some* possible world the truth of this statement (or its falsity) could be induced from a finite amount of observational material.) Yet I, for one, see no reason – not even a prima facie one – to suspect that this proposition does not have a truth value. Why *should* all truths, even all empirical truths, be discoverable by probabilistic automata (which is what I suspect we are) using a finite amount of observational material? Why does the fact that the truth value of a proposition may be undiscoverable by us suggest to some philosophers – indeed, why does it count as a *proof* for some philosophers – that the proposition in question doesn't *have* a truth value? Surely, some kind of idealistic metaphysics must be lurking in the underbrush!

What is even more startling is that philosophers who would agree with me with respect to propositions about material objects should feel differently about propositions of mathematics. (Perhaps this is due to the pernicious tendency to think of mathematics solely in terms of the mathematical-objects picture. If one doesn't understand the nature of these objects – i.e. that they don't have a 'nature', that talk about them is equivalent to talk about what is impossible – then talk about them may seem like a form of theology, and if one is anti-theological, that may be a reason for rejecting mathematics as a make-believe.) Surely, the *mere* fact that we may never know whether the continuum hypothesis is true or false is by itself just *no* reason to think that it doesn't have a truth value!

'But what does it *mean* to say that the continuum hypothesis is true?' someone will ask. It means that if S is a set of real numbers, and S is not finite and not denumerably infinite, then S can be put in one-to-one correspondence with the unit interval. Or, equivalently, it means that the sentence I have just written holds in any standard model for fourth-order number theory (actually, it can be expressed in third-order number theory). 'But what is a *standard* model?' It is one with the properties that (1) the 'integers' of the model form an ω-sequence under the $<$ of the model – i.e. it is not *possible* to select positive 'integers' $a_1, a_2, a_3, \ldots$ from the model so that, for all i, $a_{i+1} < a_i$ – and (2) the model is maximal with this property – i.e. it is not *possible* to add more 'sets' of 'integers' or 'sets of sets' of 'integers' or 'sets of sets of sets' of 'integers' to the model. (This last explanation contains the germ of the idea which is used in expressing the notion of a 'standard model' in modal-logical, as opposed to set-theoretic, language.)

I think that one can see what is going on more clearly if we imagine, for a moment, that physics has discovered that the physical universe is finite in both space and time and that all physical magnitudes are discrete (finiteness 'in the small'). That this is a possibility we must take into account was already emphasized by Hilbert in his famous article on the infinite (1926; reprinted in this volume) – it may well be, Hilbert pointed out, that we cannot argue for the consistency of any theory whose models are all infinite by arguing that physical space, or physical time, or anything else physical, provides a model for the theory, since physics is increasingly tending to replace infinities and continuities by finites and discretes.

If the whole physical universe is thoroughly finite, both in the large and in the small, then the statement '$10^{100}+1$ is a prime number' may be one whose truth value we can never know. For, if the statement is true (and even intuitionist mathematicians regard this decidable statement as possessing a truth value), then to verify that it is true by using any sieve method might well be physically impossible. And, if the shortest proof from axioms that rational beings will ever have any reason to accept is too long to be physically written out, then it might be physically impossible for beings to whom only those things are 'evident' that are in fact 'evident' (or ever will be 'evident' or that we will ever in fact have good reason to believe) to know that the statement is true.

Now, although many people doubt that the continuum hypothesis has a truth value, everyone believes that the statement '$10^{100}+1$ is a prime number' has a truth value. Why? 'Because the statement is decidable.' But what does that mean, 'the statement is decidable'? It means that it is *possible* to try out all the pairs of possible factors and see if any of them 'work'. It means that it is *possible* to decide the statement. Thus, the man who asserts that this statement is decidable, is simply making an assertion of mathematical possibility. Moreover, he believes that just one of the two statements:

> If all pairs n, m $(n, m < 10^{100}+1)$ were 'tried' by actually computing the product nm, then in some case the product would be found to equal $10^{100}+1$. (3)
>
> If all pairs $n, m, \ldots$ [same as in (3)], then in no case would the product be found to equal $10^{100}+1$. (4)

expresses a *necessary* truth, although it may be *physically* impossible to discover which one. Yet this same mathematician or philosopher, who is quite happy in this context with the notion of mathematical possibility (and who does not ask for any nominalistic reduction) and who treats mathematical necessity as well defined in this case, for a reason which is essentially circular, regards it as 'platonistic' to suppose that the continuum hypothesis has a truth value.[1] I realize that this is an ad hominem argument, but still – if there is such an intellectual sin as 'platonism' (and it is remarkably unclear what this supposed sin consists of), why is it not already to commit it, if one supposes that '$10^{100}+1$ is a prime number' has a truth value, even if no nominalistic reduction of this statement can be offered? (When one is defending a commonsense position, very often the only argument is ad hominem – for one has to keep throwing the burden of the argument back to the other side, by asking to be told *precisely* what is 'unclear' about the notions being attacked, or why a 'reduction' of the kind being demanded is necessary, or why a 'foundation' for the science in question is needed.)

In passing, I should like to remark that the following two principles, which many people seem to accept, can be shown to be inconsistent, by applying the Gödel theorem:

(I) That, even if some arithmetical (or set-theoretical) statements have no truth value, still, to say of any arithmetical (or set-theoretical) statement that it has (or lacks) a truth value is itself always either true or false (i.e. the statement either has a truth value or it doesn't).

(II) All and only the decidable statements have a truth value.

[1] Incidentally, it may also be 'platonism' to treat statements of physical possibility or counterfactual conditionals as well defined. For (1) 'physical possibility' is *compatibility* with the laws of nature. But the relation of compatibility is interdefinable with the modal notions of possibility and necessity, and, of course, the laws of nature themselves require many mathematical notions for their statement. (2) A counterfactual conditional is true just in case the consequent *follows* from the antecedent, together with certain other statements that hold both in the actual and in the hypothetical world under consideration. And, of course, no nominalistic reduction has ever succeeded, either for the notion of physical possibility or for the subjunctive conditional.

For the statement that a mathematical statement S is decidable may itself be undecidable. Then, by (II), it has no truth value to say 'S is decidable'. But, by (I), it has a truth value to say 'S has a truth value' (in fact, *falsity*; since if S has a truth value, then S is decidable, by (II), and, if S is decidable, then 'S is decidable' is also decidable). Since it is false (by the previous parenthetical remark) to say 'S has a truth value' and since we accept the equivalence of 'S has a truth value' and 'S is decidable', then it must also be *false* to say 'S is decidable'. But it has no truth value to say 'S is decidable'. Contradiction.

The significance of the antinomies

The most difficult question in the philosophy of mathematics is, perhaps, the question raised by the antinomies and by the plurality of conflicting set theories. Part of the paradox is this: the antinomies do not at all seem to affect the notion 'set of sets of integers', etc. Yet they *do* seem to affect the notion '*all* sets'. How are we to understand this situation?

One way out might be this: to conclude that we understand the notion 'set' in some contexts (e.g. 'set of integers', 'set of sets of integers'), but to conclude that we do not understand it in the context 'all sets'. But we do seem to understand *some* statements about all sets, e.g. 'for every set x and every set y, there is a set z which is the union of x and y'. So must we really abandon hope of making sense of the locution 'all sets'?

It is at this point that I urge we attend to the objects–modalities duality that I pointed out a few pages ago. The notion of a set has been used by a number of authors to clarify the notions of mathematical possibility and necessity. For example, if we identify the notion of a 'possible world' with the notion of a model (or, more correctly, with the notion of a structure of the appropriate type), then the rationale of the modal system S5 can easily be explained (as, for instance, by Carnap in *Meaning and Necessity*), and this explanation can be extended to the case of quantified modal logic by methods due to Kripke, Hintikka, and others. Here, however, I wish to go in the reverse direction, and assuming that the notions of mathematical possibility and necessity are clear (and there is no paradox associated with the notion of necessity as long as we take the '$\Box$' as a statement connective (in the degenerate sense of 'unary connective') and not – in spite of Quine's urging – as a predicate of sentences), I wish to employ these notions to try to give a clear sense to talk about 'all sets'.

My purpose is not to start a *new* school in the foundations of mathematics (say, 'modalism'). Even if in some contexts the modal-logic

picture is more helpful than the mathematical-objects picture, in other contexts the reverse is the case. Sometimes we have a clearer notion of what 'possible' means than of what 'set' means; in other cases the reverse is true; and in many, many cases both notions seem as clear as notions ever get in science. Looking at things from the standpoint of many different 'equivalent descriptions', considering what is suggested by *all* the pictures, is both a healthy antidote to foundationalism and of real heuristic value in the study of scientific questions.

Now, the natural way to interpret set-theoretic statements in the model-logical language is to interpret them as statements of what would necessarily be the case if there were standard models for the set theories in question. Since the models for von Neumann–Bernays set theory and its strengthenings (e.g. the system recently proposed by Bernays) are also models for Zermelo set theory, let me concentrate on Zermelo set theory. In order to 'concretize' the notion of a model, let us think of a model as a graph. The 'sets' of the model will then be pencil points (or some higher-dimensional analogue of pencil points, in the case of models of large cardinality), and the relation of membership will be indicated by 'arrows'. (I assume that there is nothing inconceivable about the idea of a physical space of arbitrarily high cardinality; so models of this kind need not necessarily be denumerable, and may even be standard.) Such a model will be called a 'concrete model' (or a 'standard concrete model', if it be standard) for Zermelo set theory. The model will be called standard if (1) there are no infinite-descending 'arrow' paths; and (2) it is not possible to extend the model by adding more 'sets' without adding to the number of 'ranks' in the model. (A 'rank' consists of all the sets of a given – possibly transfinite – type. 'Ranks' are cumulative types; i.e. every set of a given rank is also a set of every higher rank. It is a theorem of set theory that every set belongs to some rank.) A statement that refers only to sets of less than some given rank – say, to sets of rank less than $\omega \times 2$ – will be called a statement of 'bounded rank'. I ask the reader to accept it on faith that the statement that a certain graph G is a *standard* model for Zermelo set theory can be expressed using no 'non-nominalistic' notions except the '$\Box$'.

If S is a statement of bounded rank and if we can characterize the 'given rank' in question in some invariant way (invariant with respect to standard models of Zermelo set theory), then the statement S can easily be translated into modal-logical language. The translation is just the statement that if G is any standard model for Zermelo set theory – i.e. any standard concrete model – and G contains the invariantly characterized rank in question, then necessarily S holds in G. (It is trivial to express 'S holds in G' for any *particular* S without employing the set-theoretic notion of 'holding'.) Our problem, then, is how to translate statements of *un*bounded rank into modal-logical language.

The method is best indicated by means of an example. If the statement has the form $(x)(\exists y)(z)Mxyz$, where M is quantifier-free, then the translation is this:

If G is any standard concrete model for Zermelo set theory and if P is any point in G, then it is possible that there is a graph G' that extends G (i.e. G is a subgraph of G') and a point y in G' such that G' is a standard concrete model for Zermelo set theory and such that

(if G'' is any graph that extends G' and such that G'' is a standard concrete model for Zermelo set theory and if z is any point in G'', then $Mxyz$ holds in G'').

Obviously this method can be extended to an arbitrary set-theoretic statement.

So much for technical matters. I apologize for this brief lapse into technicality, but actually this was only the merest sketch of the technical development, and this much detail is necessary for my discussion. The real question is this: what, if any, is the philosophical significance of such translations?

If there be any philosophical significance to such translations – and I don't claim a great deal – it lies in this: I did not assume that any standard concrete model for Zermelo set theory is maximal. Indeed, I would be inclined to say that no concrete model could be maximal – nor any *non*concrete model either, as far as that goes. Even God could not make a model for Zermelo set theory that it would be *mathematically* impossible to extend, and no matter what 'stuff' He might use. Yet I succeeded in giving a clear sense to statements about 'all sets' (clear relative to the notions I assumed to start with) *without* assuming a maximal model. In metaphysical language, it is not necessary to think of sets as one system of objects in some one possible world in order to follow assertions about all sets.

Furthermore, in construing statements about sets as statements about standard concrete models for set theory, I did not introduce possible concrete models (or even possible worlds) as objects. Introducing the modal connectives '$\Box$', '$\Diamond$', '$\prec$' is not introducing new kinds of objects, but rather extending the kinds of things we can say about ordinary objects and sorts of objects. (Of course, one *can* construe the statement that it is possible that there is a graph G satisfying a condition C as meaning that *there exists a possible graph* G satisfying the condition C; that is one way of smoothing the transition from the modal-logic picture to the mathematical-objects picture.)

The importance of Zermelo set theory and of the other set theories based upon the notion of 'rank' lies in this: we have a strong intuitive conviction that whenever As are possible, so is a structure that we might call 'the family of all sets of As'. Zermelo set theory assumes only this intuition and the intuition that the process of unioning such structures can be extended into the transfinite. Of course, this intuitive conviction *may* be mistaken; it could turn out that Zermelo set theory has no standard models (even if Zermelo set theory is consistent – e.g. the discovery of an ω-inconsistency would show that there are no standard models). But so could the intuitive conviction upon which number theory is based be mistaken. If we wish to be cautious, we can assume only predicative set theory up to some 'low' transfinite type. (It is necessary to extend predicative type theory 'just past' the constructive ordinals if we wish to be able to define *validity* of schemata that contain the quantifiers 'there are infinitely many x such that' and 'there are at most a finite number of x such that', for example.) Such a weak set theory may well give us all the sets we need for physics, and also the basic notions of validity and satisfiability that we need for logic, as well as arithmetic and a weak version of classical analysis. But the fact that we do have an intuitive conviction that standard models of Zermelo set theory, or of other set theories based upon the notion of 'rank' are *mathematically possible structures* is a perfectly good reason for asking what statements necessarily hold in such structures – e.g. for asking whether the continuum hypothesis necessarily holds in such structures.

The real significance of the Russell paradox, from the standpoint of the modal-logic picture, is this: it shows that *no* concrete structure can be a standard model for the naive conception of the totality of all sets; for any concrete structure has a possible extension that contains more 'sets'. (If we identify sets with the points that represent them in the various possible concrete structures, we might say: it is not possible for all *possible* sets to exist in any one world!) Yet set theory does not become impossible. Rather, set theory becomes the study of what must hold in, e.g. any standard model for Zermelo set theory.