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.

PART III

Mathematical truth

The *a priori*

ALFRED JULES AYER

The view of philosophy which we have adopted may, I think, fairly be described as a form of empiricism. For it is characteristic of an empiricist to eschew metaphysics, on the ground that every factual proposition must refer to sense-experience. And even if the conception of philosophizing as an activity of analysis is not to be discovered in the traditional theories of empiricists, we have seen that it is implicit in their practice. At the same time, it must be made clear that, in calling ourselves empiricists, we are not avowing a belief in any of the psychological doctrines which are commonly associated with empiricism. For, even if these doctrines were valid, their validity would be independent of the validity of any philosophical thesis. It could be established only by observation, and not by the purely logical considerations upon which our empiricism rests.

Having admitted that we are empiricists, we must now deal with the objection that is commonly brought against all forms of empiricism; the objection, namely, that it is impossible on empiricist principles to account for our knowledge of necessary truths. For, as Hume conclusively showed, no general proposition whose validity is subject to the test of actual experience can ever be logically certain. No matter how often it is verified in practice, there still remains the possibility that it will be confuted on some future occasion. The fact that a law has been substantiated in $n-1$ cases affords no logical guarantee that it will be substantiated in the nth case also, no matter how large we take n to be. And this means that no general proposition referring to a matter of fact can ever be shown to be necessarily and universally true. It can at best be a probable hypothesis. And this, we shall find, applies not only to general propositions, but to all propositions which have a factual content. They can none of them ever become logically certain. This conclusion, which we shall elaborate later on, is one which must be accepted by every consistent empiricist. It is often thought to involve him in complete scepticism; but this is not the case. For the fact that the validity of a proposition cannot be logically guaranteed in no way entails that it is irrational

Excerpted and reprinted with the kind permission of the author and publishers from Alfred Jules Ayer, *Language, Truth and Logic* (London: Victor Gollancz, Ltd., 1956; New York: Dover Publications, Inc.), pp. 71-87.

for us to believe it. On the contrary, what is irrational is to look for a guarantee where none can be forthcoming; to demand certainty where probability is all that is obtainable. We have already remarked upon this, in referring to the work of Hume. And we shall make the point clearer when we come to treat of probability, in explaining the use which we make of empirical propositions. We shall discover that there is nothing perverse or paradoxical about the view that all the "truths" of science and common sense are hypotheses; and consequently that the fact that it involves this view constitutes no objection to the empiricist thesis.

Where the empiricist does encounter difficulty is in connection with the truths of formal logic and mathematics. For whereas a scientific generalization is readily admitted to be fallible, the truths of mathematics and logic appear to everyone to be necessary and certain. But if empiricism is correct no proposition which has a factual content can be necessary or certain. Accordingly the empiricists must deal with the truths of logic and mathematics in one of the two following ways: he must say either that they are not necessary truths, in which case he must account for the universal conviction that they are; or he must say that they have no factual content, and then he must explain how a proposition which is empty of all factual content can be true and useful and surprising.

If neither of these courses proves satisfactory, we shall be obliged to give way to rationalism. We shall be obliged to admit that there are some truths about the world which we can know independently of experience; that there are some properties which we can ascribe to all objects, even though we cannot conceivably observe that all objects have them. And we shall have to accept it as a mysterious inexplicable fact that our thought has this power to reveal to us authoritatively the nature of objects which we have never observed. Or else we must accept the Kantian explanation which, apart from the epistemological difficulties which we have already touched on, only pushes the mystery a stage further back.

It is clear that any such concession to rationalism would upset the main argument of this book. For the admission that there were some facts about the world which could be known independently of experience would be incompatible with our fundamental contention that a sentence says nothing unless it is empirically verifiable. And thus the whole force of our attack on metaphysics would be destroyed. It is vital, therefore, for us to be able to show that one or other of the empiricist accounts of the propositions of logic and mathematics is correct. If we are successful in this, we shall have destroyed the foundations of rationalism. For the fundamental tenet of rationalism is that thought is an independent source of knowledge, and is moreover a more trustworthy source of knowledge

than experience; indeed some rationalists have gone so far as to say that thought is the only source of knowledge. And the ground for this view is simply that the only necessary truths about the world which are known to us are known through thought and not through experience. So that if we can show either that the truths in question are not necessary or that they are not "truths about the world," we shall be taking away the support on which rationalism rests. We shall be making good the empiricist contention that there are no "truths of reason" which refer to matters of fact.

The course of maintaining that the truths of logic and mathematics are not necessary or certain was adopted by Mill. He maintained that these propositions were inductive generalizations based on an extremely large number of instances. The fact that the number of supporting instances was so very large accounted, in his view, for our believing these generalizations to be necessarily and universally true. The evidence in their favor was so strong that it seemed incredible to us that a contrary instance should ever arise. Nevertheless it was in principle possible for such generalizations to be confuted. They were highly probable, but, being inductive generalizations, they were not certain. The difference between them and the hypotheses of natural science was a difference in degree and not in kind. Experience gave us very good reason to suppose that a "truth" of mathematics or logic was true universally; but we were not possessed of a guarantee. For these "truths" were only empirical hypotheses which had worked particularly well in the past; and, like all empirical hypotheses, they were theoretically fallible.

I do not think that this solution of the empiricist's difficulty with regard to the propositions of logic and mathematics is acceptable. In discussing it, it is necessary to make a distinction which is perhaps already enshrined in Kant's famous dictum that, although there can be no doubt that all our knowledge begins with experience, it does not follow that it all arises out of experience (Kant 1881: Introduction, section i). When we say that the truths of logic are known independently of experience, we are not of course saying that they are innate, in the sense that we are born knowing them. It is obvious that mathematics and logic have to be learned in the same way as chemistry and history have to be learned. Nor are we denying that the first person to discover a given logical or mathematical truth was led to it by an inductive procedure. It is very probable, for example, that the principle of the syllogism was formulated not before but after the validity of syllogistic reasoning had been observed in a number of particular cases. What we are discussing, however, when we say that logical and mathematical truths are known independently of experience, is not a historical question concerning the way in which these truths were originally discovered, not a psychological question concerning the way in which each of us comes to learn them, but an epistemological question. The contention of Mill's which we reject is that the propositions of logic and mathematics have the same status as empirical hypotheses; that their validity is determined in the same way. We maintain that they are independent of experience in the sense that they do not owe their validity to empirical verification. We may come to discover them through an inductive process; but once we have apprehended them we see that they are necessarily true, that they hold good for every conceivable instance. And this serves to distinguish them from empirical generalizations. For we know that a proposition whose validity depends upon experience cannot be seen to be necessarily and universally true.

In rejecting Mill's theory, we are obliged to be somewhat dogmatic. We can do no more than state the issue clearly and then trust that his contention will be seen to be discrepant with the relevant logical facts. The following considerations may serve to show that of the two ways of dealing with logic and mathematics which are open to the empiricist, the one which Mill adopted is not the one which is correct.

The best way to substantiate our assertion that the truths of formal logic and pure mathematics are necessarily true is to examine cases in which they might seem to be confuted. It might easily happen, for example, that when I came to count what I had taken to be five pairs of objects, I found that they amounted only to nine. And if I wished to mislead people I might say that on this occasion twice five was not ten. But in that case I should not be using the complex sign "$2 \times 5 = 10$" in the way in which it is ordinarily used. I should be taking it not as the expression of a purely mathematical proposition, but as the expression of an empirical generalization, to the effect that whenever I counted what appeared to be to be five pairs of objects I discovered that they were ten in number. This generalization may very well be false. But if it proved false in a given case, one would not say that the mathematical proposition "$2 \times 5 = 10$" had been confuted. One would say that I was wrong in supposing that there were five pairs of objects to start with, or that one of the objects had been taken away while I was counting, or that two of them had coalesced, or that I had counted wrongly. One would adopt as an explanation whatever empirical hypothesis fitted in best with the accredited facts. The one explanation which would in no circumstances be adopted is that ten is not always the product of two and five.

To take another example: if what appears to be a Euclidean triangle is found by measurement not to have angles totalling 180 degrees, we do not say that we have met with an instance which invalidates the mathematical proposition that the sum of the three angles of a Euclidean triangle is 180 degrees. We say that we have measured wrongly, or, more probably, that the triangle we have been measuring is not Euclidean. And this is our procedure in every case in which a mathematical truth might appear to be confuted. We always preserve its validity by adopting some other explanation of the occurrence.

The same thing applies to the principles of formal logic. We may take an example relating to the so-called law of excluded middle, which states that a proposition must be either true or false, or, in other words, that it is impossible that a proposition and its contradictory should neither of them be true. One might suppose that a proposition of the form "*x* has stopped doing *y*" would in certain cases constitute an exception to this law. For instance, if my friend has never yet written to me, it seems fair to say that it is neither true nor false that he has stopped writing to me. But in fact one would refuse to accept such an instance as an invalidation of the law of excluded middle. One would point out that the proposition "My friend has stopped writing to me" is not a simple proposition, but the conjunction of the two propositions "My friend wrote to me in the past" and "My friend does not write to me now": and, furthermore, that the proposition "My friend has not stopped writing to me" is not, as it appears to be, contradictory to "My friend has stopped writing to me," but only contrary to it. For it means "My friend wrote to me in the past, and he still writes to me." When, therefore, we say that such a proposition as "My friend has stopped writing to me" is sometimes neither true nor false, we are speaking inaccurately. For we seem to be saying that neither it nor its contradictory is true. Whereas what we mean, or anyhow should mean, is that neither it nor its apparent contradictory is true. And its apparent contradictory is really only its contrary. Thus we preserve the law of excluded middle by showing that the negating of a sentence does not always yield the contradictory of the proposition originally expressed.

There is no need to give further examples. Whatever instance we care to take, we shall always find that the situations in which a logical or mathematical principle might appear to be confuted are accounted for in such a way as to leave the principle unassailed. And this indicates that Mill was wrong in supposing that a situation could arise which would overthrow a mathematical truth. The principles of logic and mathematics are true universally simply because we never allow them to be anything else. And the reason for this is that we cannot abandon them without contradicting ourselves, without sinning against the rules which govern the use of language, and so making our utterances self-stultifying. In other words, the truths of logic and mathematics are analytic propositions or tautologies. In saying this we are making what will be held to be an extremely controversial statement, and we must now proceed to make its implications clear.

The most familiar definition of an analytic proposition, or judgment, as he called it, is that given by Kant. He said that an analytic judgment was one in which the predicate B belonged to the subject A as something which was covertly contained in the concept of A (Kant 1881: Introduction, sections iv and v). He contrasted analytic with synthetic judgments, in which the predicate B lay outside the subject A, although it did stand in connection with it. Analytic judgments, he explains, "add nothing through the predicate to the concept of the subject, but merely break it up into those constituent concepts that have all along been thought in it, although confusedly." Synthetic judgments, on the other hand, "add to the concept of the subject a predicate which has not been in any wise thought in it, and which no analysis could possibly extract from it." Kant gives "all bodies are extended" as an example of an analytic judgment, on the ground that the required predicate can be extracted from the concept of "body," "in accordance with the principle of contradiction"; as an example of a synthetic judgment, he gives "all bodies are heavy." He refers also to "7+5=12" as a synthetic judgment, on the ground that the concept of twelve is by no means already thought in merely thinking the union of seven and five. And he appears to regard this as tantamount to saying that the judgment does not rest on the principle of contradiction alone. He holds, also, that through analytic judgments our knowledge is not extended as it is through synthetic judgments. For in analytic judgments "the concept which I already have is merely set forth and made intelligible to me."

I think that this is a fair summary of Kant's account of the distinction between analytic and synthetic propositions, but I do not think that it succeeds in making the distinction clear. For even if we pass over the difficulties which arise out of the use of the vague term "concept," and the unwarranted assumption that every judgment, as well as every German or English sentence, can be said to have a subject and a predicate, there remains still this crucial defect. Kant does not give one straightforward criterion for distinguishing between analytic and synthetic propositions; he gives two distinct criteria, which are by no means equivalent. Thus his ground for holding that the proposition "7+5=12" is synthetic is, as we have seen, that the subjective intension of "7+5" does not comprise the subjective intension of "12"; whereas his ground for holding that "all bodies are extended" is an analytic proposition is that it rests on the principle of contradiction alone. That is, he employs a psychological criterion in the first of these examples, and a logical criterion in the second, and takes their equivalence for granted. But, in fact,

a proposition which is synthetic according to the former criterion may very well be analytic according to the latter. For, as we have already pointed out, it is possible for symbols to be synonymous without having the same intentional meaning for anyone: and accordingly from the fact that one can think of the sum of seven and five without necessarily thinking of twelve, it by no means follows that the proposition "7+5=12" can be denied without self-contradiction. From the rest of his argument, it is clear that it is this logical proposition, and not any psychological proposition, that Kant is really anxious to establish. His use of the psychological criterion leads him to think that he has established it, when he has not.

I think that we can preserve the logical import of Kant's distinction between analytic and synthetic propositions, while avoiding the confusions which mar his actual account of it, if we say that a proposition is analytic when its validity depends solely on the definitions of the symbols it contains, and synthetic when its validity is determined by the facts of experience. Thus, the proposition "There are ants which have established a system of slavery" is a synthetic proposition. For we cannot tell whether it is true or false merely by considering the definitions of the symbols which constitute it. We have to resort to actual observation of the behaviour of ants. On the other hand, the proposition "Either some ants are parasitic or none are" is an analytic proposition. For one need not resort to observation to discover that there either are or are not ants which are parasitic. If one knows what is the function of the words "either," "or," and "not," then one can see that any proposition of the form "Either *p* is true or *p* is not true" is valid, independently of experience. Accordingly, all such propositions are analytic.

It is to be noticed that the proposition "Either some ants are parasitic or none are" provides no information whatsoever about the behavior of ants, or, indeed, about any matter of fact. And this applies to all analytic propositions. They none of them provide any information about any matter of fact. In other words, they are entirely devoid of factual content. And it is for this reason that no experience can confute them.

When we say that analytic propositions are devoid of factual content; and consequently that they say nothing, we are not suggesting that they are senseless in the way that metaphysical utterances are senseless. For, although they give us no information about any empirical situation, they do enlighten us by illustrating the way in which we use certain symbols. Thus if I say, "Nothing can be colored in different ways at the same time with respect to the same part of itself," I am not saying anything about the properties of any actual thing; but I am not talking nonsense. I am expressing an analytic proposition, which records our determination to call a color expanse which differs in quality from a neighboring color expanse a different part of a given thing. In other words, I am simply calling attention to the implications of a certain linguistic usage. Similarly, in saying that if all Bretons are Frenchmen, and all Frenchmen Europeans, then all Bretons are Europeans, I am not describing any matter of fact. But I am showing that in the statement that all Bretons are Frenchmen, and all Frenchmen Europeans, the further statement that all Bretons are Europeans is implicitly contained. And I am thereby indicating the convention which governs our usage of the words "if" and "all."

We see, then, that there is a sense in which analytic propositions do give us new knowledge. They call attention to linguistic usages, of which we might otherwise not be conscious, and they reveal unsuspected implications in our assertions and beliefs. But we can see also that there is a sense in which they may be said to add nothing to our knowledge. For they tell us only what we may be said to know already. Thus, if I know that the existence of May Queens is a relic of tree-worship, and I discover that May Queens still exist in England, I can employ the tautology "If *p* implies *q*, and *p* is true, *q* is true" to show that there still exists a relic of tree-worship in England. But in saying that there are still May Queens in England, and that the existence of May Queens is a relic of tree-worship, I have already asserted the existence in England of a relic of tree-worship. The use of the tautology does, indeed, enable me to make this concealed assertion explicit. But it does not provide me with any new knowledge, in the sense in which empirical evidence that the election of May Queens had been forbidden by law would provide me with new knowledge. If one had to set forth all the information one possessed, with regard to matters of fact, one would not write down any analytic propositions. But one would make use of analytic propositions in compiling one's encyclopaedia, and would thus come to include propositions which one would otherwise have overlooked. And, besides enabling one to make one's list of information complete, the formulation of analytic propositions would enable one to make sure that the synthetic propositions of which the list was composed formed a self-consistent system. By showing which ways of combining propositions resulted in contradictions, they would prevent one from including incompatible propositions and so making the list self-stultifying. But insofar as we had actually used words as "all" and "or" and "not" without falling into self-contradiction, we might be said already to know what was revealed in the formulation of analytic propositions illustrating the rules which govern our usage of these logical particles. So that here again we are justified in saying that analytic propositions do not increase our knowledge.

The analytic character of the truths of formal logic was obscured in the traditional logic through its being insufficiently formalized. For in speaking always of judgments, instead of propositions, and introducing irrelevant psychological questions, the traditional logic gave the impression of being concerned in some specially intimate way with the workings of thought. What it was actually concerned with was the formal relationship of classes, as is shown by the fact that all its principles of inference are subsumed in the Boolean class-calculus, which is subsumed in its turn in the propositional calculus of Russell and Whitehead (cf. Menger 1933: 94–6; and Lewis and Langford 1932, chap. 5). Their system, expounded in *Principia Mathematica,* makes it clear that formal logic is not concerned with the properties of men's minds, much less with the properties of material objects, but simply with the possibility of combining propositions by means of logical particles into analytic propositions, and with studying the formal relationship of these analytic propositions, in virtue of which one is deducible from another. Their procedure is to exhibit the propositions of formal logic as a deductive system, based on five primitive propositions, subsequently reduced in number to one. Hereby the distinction between logical truths and principles of inference, which was maintained in the Aristotelian logic, very properly disappears. Every principle of inference is put forward as a logical truth and every logical truth can serve as a principle of inference. The three Aristotelian "laws of thought," the law of identity, the law of excluded middle, and the law of non-contradiction, are incorporated in the system, but they are not considered more important than the other analytic propositions. They are not reckoned among the premises of the system. And the system of Russell and Whitehead itself is probably only one among many possible logics, each of which is composed of tautologies as interesting to the logician as the arbitrarily selected Aristotelian "laws of thought" (cf. Lewis and Langford 1932, chap. 7).

A point which is not sufficiently brought out by Russell, if indeed it is recognized by him at all, is that every logical proposition is valid in its own right. Its validity does not depend on its being incorporated in a system, and deduced from certain propositions which are taken as self-evident. The construction of systems of logic is useful as a means of discovering and certifying analytic propositions, but it is not in principle essential even for this purpose. For it is possible to conceive of a symbolism in which every analytic proposition could be seen to be analytic in virtue of its form alone.

The fact that the validity of an analytic proposition in no way depends on its being deducible from other analytic propositions is our justification for disregarding the question whether the propositions of mathematics are reducible to propositions of formal logic, in the way that Russell supposed (1919, chap. 2) [pp. 167–73 in this volume]. For even if it is the case that the definition of a cardinal number as a class of classes similar to a given class is circular, and it is not possible to reduce mathematical notions to purely logical notions, it will still remain true that the propositions of mathematics are analytic propositions. They will form a special class of analytic propositions, containing special terms, but they will be none the less analytic for that. For the criterion of an analytic proposition is that its validity should follow simply from the definition of the terms contained in it, and this condition is fulfilled by the propositions of pure mathematics.

The mathematical propositions which one might most pardonably suppose to be synthetic are the propositions of geometry. For it is natural for us to think, as Kant thought, that geometry is the study of the properties of physical space, and consequently that its propositions have factual content. And if we believe this, and also recognize that the truths of geometry are necessary and certain, then we may be inclined to accept Kant's hypothesis that space is the form of intuition of our outer sense, a form imposed by us on the matter of sensation, as the only possible explanation of our *a priori* knowledge of these synthetic propositions. But while the view that pure geometry is concerned with physical space was plausible enough in Kant's day, when the geometry of Euclid was the only geometry known, the subsequent invention of non-Euclidean geometries has shown it to be mistaken. We see now that the axioms of a geometry are simply definitions, and that the theorems of a geometry are simply the logical consequences of these definitions. A geometry is not in itself about physical space; in itself it cannot be said to be "about" anything. But we can use a geometry to reason about physical space. That is to say, once we have given the axioms a physical interpretation, we can proceed to apply the theorems to the objects which satisfy the axioms (cf. Poincaré 1903: pt. 2, chap. 3). Whether a geometry can be applied to the actual physical world or not, is an empirical question which falls outside the scope of the geometry itself. There is no sense, therefore, in asking which of the various geometries known to us are false, and which are true. Insofar as they are all free from contradiction, they are all true. What one can ask is which of them is the most useful on any given occasion, which of them can be applied most easily and most fruitfully to an actual empirical situation. But the proposition which states that a certain application of a geometry is possible is not itself a proposition of that geometry. All that the geometry itself tells us is that if anything can be brought under the definitions, it will also satisfy the theorems. It is therefore a purely logical system, and its propositions are purely analytic propositions.

It might be objected that the use made of diagrams in geometrical treatises shows that geometrical reasoning is not purely abstract and logical, but depends on our intuition of the properties of figures. In fact, however, the use of diagrams is not essential to completely rigorous geometry. The diagrams are introduced as an aid to our reason. They provide us with a particular application of the geometry, and so assist us to perceive the more general truth that the axioms of the geometry involve certain consequences. But the fact that most of us need the help of an example to make us aware of those consequences does not show that the relation between them and the axioms is not a purely logical relation. It shows merely that our intellects are unequal to the task of carrying out very abstract processes of reasoning without the assistance of intuition. In other words, it has no bearing on the nature of geometrical propositions, but is simply an empirical fact about ourselves. Moreover, the appeal to intuition, though generally of psychological value, is also a source of danger to the geometer. He is tempted to make assumptions which are accidentally true of the particular figure he is taking as an illustration, but do not follow from his axioms. It has, indeed, been shown that Euclid himself was guilty of this, and consequently that the presence of the figure is essential to some of his proofs (cf. Black 1933: 154). This shows that his system is not, as he presents it, completely rigorous, although of course it can be made so. It does not show that the presence of the figure is essential to a truly rigorous geometrical proof. To suppose that it did would be to take as a necessary feature of all geometries what is really only an incidental defect in one particular geometrical system.

We conclude, then, that the propositions of pure geometry are analytic. And this leads us to reject Kant's hypothesis that geometry deals with the form of intuition of our outer sense. For the ground for this hypothesis was that it alone explained how the propositions of geometry could be both true *a priori* and synthetic: and we have seen that they are not synthetic. Similarly our view that the propositions of arithmetic are not synthetic but analytic leads us to reject the Kantian hypothesis[1] that arithmetic is concerned with our pure intuition of time, the form of our inner sense. And thus we are able to dismiss Kant's transcendental aesthetic without having to bring forward the epistemological difficulties which it is commonly said to involve. For the only argument which can

[1]This hypothesis is not mentioned in the *Critique of Pure Reason*, but was maintained by Kant at an earlier date.

be brought in favour of Kant's theory is that it alone explains certain "facts." And now we have found that the "facts" which it purports to explain are not facts at all. For while it is true that we have *a priori* knowledge of necessary propositions, it is not true, as Kant supposed, that any of these necessary propositions are synthetic. They are without exception analytic propositions, or, in other words, tautologies.

We have already explained how it is that these analytic propositions are necessary and certain. We saw that the reason why they cannot be confuted in experience is that they do not make any assertion about the empirical world. They simply record our determination to use words in a certain fashion. We cannot deny them without infringing the conventions which are presupposed by our very denial, and so falling into self-contradiction. And this is the sole ground of their necessity. As Wittgenstein puts it, our justification for holding that the world could not conceivably disobey the laws of logic is simply that we could not say of an unlogical world how it would look (1922: 3.01). And just as the validity of an analytic proposition is independent of the nature of the external world; so is it independent of the nature of our minds. It is perfectly conceivable that we should have employed different linguistic conventions from those which we actually do employ. But whatever these conventions might be, the tautologies in which we recorded them would always be necessary. For any denial of them would be self-stultifying.

We see, then, that there is nothing mysterious about the apodeictic certainty of logic and mathematics. Our knowledge that no observation can ever confute the proposition "$7+5=12$" depends simply on the fact that the symbolic expression "$7+5$" is synonymous with "12," just as our knowledge that every oculist is an eye-doctor depends on the fact that the symbol "eye-doctor" is synonymous with "oculist." And the same explanation holds good for every other *a priori* truth.

What is mysterious at first sight is that these tautologies should on occasion be so surprising, that there should be in mathematics and logic the possibility of invention and discovery. As Poincaré says: "If all the assertions which mathematics puts forward can be derived from one another by formal logic, mathematics cannot amount to anything more than an immense tautology. Logical inference can teach us nothing essentially new, and if everything is to proceed from the principle of identity, everything must be reducible to it. But can we really allow that these theorems which fill so many books serve no other purpose than to say in a round-about fashion '$A=A$'?" (Poincaré 1903: pt. 1, chap. 1). Poincaré finds this incredible. His own theory is that the sense of invention and discovery in mathematics belongs to it in virtue of mathematical induction, the principle that what is true for the number 1, and true for

$n+1$ when it is true for n,[2] is true for all numbers. And he claims that this is a synthetic *a priori* principle. It is, in fact, *a priori*, but it is not synthetic. It is a defining principle of the natural numbers, serving to distinguish them from such numbers as the infinite cardinal numbers, to which it cannot be applied (cf. Russell 1919: 27). Moreover, we must remember that discoveries can be made, not only in arithmetic, but also in geometry and formal logic, where no use is made of mathematical induction. So that even if Poincaré were right about mathematical induction, he would not have provided a satisfactory explanation of the paradox that a mere body of tautologies can be so interesting and so surprising.

The true explanation is very simple. The power of logic and mathematics to surprise us depends, like their usefulness, on the limitations of our reason. A being whose intellect was infinitely powerful would take no interest in logic and mathematics (cf. Hahn 1933: 18). For he would be able to see at a glance everything that his definitions implied, and, accordingly, could never learn anything from logical inference which he was not fully conscious of already. But our intellects are not of this order. It is only a minute proportion of the consequences of our definitions that we are able to detect at a glance. Even so simple a tautology as "$91\times79=7189$" is beyond the scope of our immediate apprehension. To assure ourselves that "7189" is synonymous with "91×79" we have to resort to calculation, which is simply a process of tautological transformation – that is, a process by which we change the form of expressions without altering their significance. The multiplication tables are rules for carrying out this process in arithmetic, just as the laws of logic are rules for the tautological transformation of sentences expressed in logical symbolism or in ordinary language. As the process of calculation is carried out more or less mechanically, it is easy for us to make a slip and so unwittingly contradict ourselves. And this accounts for the existence of logical and mathematical "falsehoods," which otherwise might appear paradoxical. Clearly the risk of error in logical reasoning is proportionate to the length and the complexity of the process of calculation. And in the same way, the more complex an analytic proposition is, the more chance it has of interesting and surprising us.

It is easy to see that the danger of error in logical reasoning can be minimized by the introduction of symbolic devices, which enable us to express highly complex tautologies in a conveniently simple form. And this gives us an opportunity for the exercise of invention in the pursuit of logical enquiries. For a well-chosen definition will call our attention to

[2] This was wrongly stated in previous editions as "true for n when it is true for $n+1$."

analytic truths, which would otherwise have escaped us. And the framing of definitions which are useful and fruitful may well be regarded as a creative act.

Having thus shown that there is no inexplicable paradox involved in the view that the truths of logic and mathematics are all of them analytic, we may safely adopt it as the only satisfactory explanation of their *a priori* necessity. And in adopting it we vindicate the empiricist claim that there can be no *a priori* knowledge of reality. For we show that the truths of pure reason, the proportions which we know to be valid independently of all experience, are so only in virtue of their lack of factual content. To say that a proposition is true *a priori* is to say that it is a tautology. And tautologies, though they may serve to guide us in our empirical search for knowledge, do not in themselves contain any information about any matter of fact.

Truth by convention

W. V. QUINE

The less a science has advanced, the more its terminology tends to rest on an uncritical assumption of mutual understanding. With increase of rigor this basis is replaced piecemeal by the introduction of definitions. The interrelationships recruited for these definitions gain the status of analytic principles; what was once regarded as a theory about the world becomes reconstrued as a convention of language. Thus it is that some flow from the theoretical to the conventional is an adjunct of progress in the logical foundations of any science. The concept of simultaneity at a distance affords a stock example of such development: in supplanting the uncritical use of this phrase by a definition. Einstein so chose the definitive relationship as to verify conventionally the previously paradoxical principle of the absoluteness of the speed of light. But whereas the physical sciences are generally recognized as capable only of incomplete evolution in this direction, and as destined to retain always a nonconventional kernel of doctrine, developments of the past few decades have led to a widespread conviction that logic and mathematics are purely analytic or conventional. It is less the purpose of the present inquiry to question the validity of this contrast than to question its sense.

I

A definition, strictly, is a convention of notational abbreviation (cf. Russell 1903: 429). A *simple* definition introduces some specific expression, e.g., 'kilometer', or '*e*', called the *definiendum*, as arbitrary shorthand for some complex expression, e.g., 'a thousand meters' or '$\lim_{n\to\infty}(1+1/n)^n$', called the *definiens*. A *contextual* definition sets up indefinitely many mutually analogous pairs of definienda and definientia according to some general scheme; an example is the definition whereby expressions of the form 'sin ——/cos ——' are abbreviated as 'tan ——'. From a formal standpoint the signs thus introduced are

First published in Otis H. Lee, ed., *Philosophical Essays for A. N. Whitehead* (New York: Longmans, Green & Company, Inc., 1936). Reprinted with the kind permission of the author and David McKay Company. Some minor corrections have been made by the author, particularly on p. 339.

wholly arbitrary; all that is required of a definition is that it be theoretically immaterial, i.e., that the shorthand which it introduces admit in every case of unambiguous elimination in favor of the antecedent longhand.[1]

Functionally a definition is not a premiss to theory, but a license for rewriting theory by putting definiens for definiendum or vice versa. By allowing such replacements a definition transmits truth: it allows true statements to be translated into new statements which are true by the same token. Given the truth of the statement 'The altitude of Kibo exceeds six thousand meters', the definition of 'kilometer' makes for the truth of the statement 'The altitude of Kibo exceeds six kilometers'; given the truth of the statement '$\sin \pi/\cos \pi = \sin \pi/\cos \pi$,' of which logic assures us in its earliest pages, the contextual definition cited above makes for the truth of the statement '$\tan \pi = \sin j/\cos j$.' In each case the statement inferred through the definition is true only because it is shorthand for another statement which was true independently of the definition. Considered in isolation from all doctrine, including logic, a definition is incapable of grounding the most trivial statement; even '$\tan \pi = \sin \pi/\cos \pi$' is a definitional transformation of an antecedent self-identity, rather than a spontaneous consequence of the definition.

What is loosely called a logical consequence of definitions is therefore more exactly describable as a logical truth definitionally abbreviated: a statement which becomes a truth of logic when definienda are replaced by definientia. In this sense '$\tan \pi = \sin \pi/\cos \pi$' is a logical consequence of the contextual definition of the tangent. 'The altitude of Kibo exceeds six kilometers' is not *ipso facto* a logical consequence of the given definition of 'kilometer'; on the other hand it would be a logical consequence of a quite suitable but unlikely definition introducing 'Kibo' as an abbreviation of the phrase 'the totality of such African terrain as exceeds six kilometers in altitude', for under this definition the statement in question is an abbreviation of a truth of logic, viz., 'The altitude of the totality of such African terrain as exceeds six kilometers in altitude exceeds six kilometers.'

Whatever may be agreed upon as the exact scope of logic, we may expect definitional abbreviations of logical truths to be reckoned as logical rather than extralogical truths. This being the case, the preceding

[1]From the present point of view a contextual definition may be recursive, but can then count among its definienda only those expressions in which the argument of recursion has a constant value, since otherwise the requirement of eliminability is violated. Such considerations are of little consequence, however, since any recursive definition can be turned into a direct one by purely logical methods. Cf. Carnap 1934b: 23, 79.

conclusion shows logical consequences of definitions to be themselves truths of logic. To claim that mathematical truths are conventional in the sense of following logically from definitions is therefore to claim that mathematics is part of logic. The latter claim does not represent an arbitrary extension of the term 'logic' to include mathematics; agreement as to what belongs to logic and what belongs to mathematics is supposed at the outset, and it is then claimed that definitions of mathematical expressions can so be framed on the basis of logical ones that all mathematical truths become abbreviations of logical ones.

Although signs introduced by definition are formally arbitrary, more than such arbitrary notational convention is involved in questions of definability; otherwise any expression might be said to be definable on the basis of any expressions whatever. When we speak of definability, or of finding a definition for a given sign, we have in mind some traditional usage of the sign antecedent to the definition in question. To be satisfactory in this sense a definition of the sign not only must fulfill the formal requirement of unambiguous eliminability, but must also conform to the traditional usage in question. For such conformity it is necessary and sufficient that every context of the sign which was true and every context which was false under traditional usage be construed by the definition as an abbreviation of some other statement which is correspondingly true or false under the established meanings of its signs. Thus when definitions of mathematical expressions on the basis of logical ones are said to have been framed, what is meant is that definitions have been set up whereby every statement which so involves those mathematical expressions as to be recognized traditionally as true, or as false, is construed as an abbreviation of another correspondingly true or false statement which lacks those mathematical expressions and exhibits only logical expressions in their stead.[2]

An expression will be said to occur *vacuously* in a given statement if its replacement therein by any and every other grammatically admissible expression leaves the truth or falsehood of the statement unchanged. Thus for any statement containing some expressions vacuously there is a class of statements, describable as *vacuous variants* of the given statement, which are like it in point of truth or falsehood, like it also in point of a certain skeleton of symbolic make-up, but diverse in exhibiting all grammatically possible variations upon the vacuous constituents of the

[2]Note than an expression is said to be defined, in terms, e.g., of logic, not only when it is a single sign whose elimination from a context in favor of logical expressions is accomplished by a single application of one definition, but also when it is a complex expression whose elimination calls for successive application of many definitions.

given statement. An expression will be said to occur *essentially* in a statement if it occurs in all the vacuous variants of the statement, i.e., if it forms part of the aforementioned skeleton. (Note that though an expression occurs non-vacuously in a statement it may fail of essential occurrence because some of its parts occur vacuously in the statement.)

Now let S be a truth, let the expressions E_i occur vacuously in S, and let the statements S_i be the vacuous variants of S. Thus the S_i will likewise be true. On the sole basis of the expressions belonging to a certain class α, let us frame a definition for one of the expressions F occurring in S outside the E_i. S and the S_i thereby become abbreviations of certain statements S' and S_i' which exhibit only members of α instead of those occurrences of F, but which remain so related that the S_i' are all the results of replacing the E_i in S_i' by any other grammatically admissible expressions. Now since our definition of F is supposed to conform to usage, S' and the S_i' will, like S and the S_i, be uniformly true; hence the S_i' will be vacuous variants of S', and the occurrences of the E_i in S' will be vacuous. The definition thus makes S an abbreviation of a truth S' which, like S, involves the E_i vacuously, but which differs from S in exhibiting only members of α instead of the occurrences of F outside the E_i. Now it is obvious that an expression cannot occur essentially in a statement if it occurs only within expressions which occur vacuously in the statement; consequently F, occurring in S' as it does only within the E_i if at all, does not occur essentially in S'; members of α occur essentially in its stead. Thus if we take F as any non-member of α occurring essentially in S, and repeat the above reasoning for each such expression, we see that, through definitions of all such expressions in terms of members of α, S becomes an abbreviation of a truth S'' involving only members of α essentially.

Thus if in particular we take α as the class of all logical expressions, the above tells us that if logical definitions be framed for all non-logical expressions occurring essentially in the true statement S, S becomes an abbreviation of a truth S'' involving only logical expressions essentially. But if S'' involves only logical expressions essentially, and hence remains true when everything except that skeleton of logical expressions is changed in all grammatically possible ways, then S'' depends for its truth upon those logical constituents alone, and is thus a truth of logic. It is therefore established that if all non-logical expressions occurring essentially in a true statement S be given definitions on the basis solely of logic, then S becomes an abbreviation of a truth S'' of logic. In particular, then, if all mathematical expressions be defined in terms of logic, all truths involving only mathematical and logical expressions essentially become definitional abbreviations of truths of logic.

Now a mathematical truth, e.g., 'Smith's age plus Brown's equals Brown's age plus Smith's,' may contain non-logical, non-mathematical expressions. Still any such mathematical truth, or another whereof it is a definitional abbreviation, will consist of a skeleton of mathematical or logical expressions filled in with non-logical, non-mathematical expressions all of which occur vacuously. Thus every mathematical truth either is a truth in which only mathematical and logical expressions occur essentially, or is a definitional abbreviation of such a truth. Hence, granted definitions of all mathematical expressions in terms of logic, the preceding conclusion shows that all mathematical truths become definitional abbreviations of truths of logic – therefore truths of logic in turn. For the thesis that mathematics is logic it is thus sufficient that all mathematical notation be defined on the basis of logical notation.

If on the other hand some mathematical expressions resist definition on the basis of logical ones, then every mathematical truth containing such recalcitrant expressions must contain them only inessentially, or be a definitional abbreviation of a truth containing such expressions only inessentially, if all mathematics is to be logic: for though a logical truth, e.g., the above one about Africa, may involve non-logical expressions, it or some other logical truth whereof it is an abbreviation must involve only logical expressions essentially. It is of this alternative that those avail themselves who regard mathematical truths, insofar as they depend upon non-logical notions, as elliptical for hypothetical statements containing as tacit hypotheses all the postulates of the branch of mathematics in question (e.g., Russell 1903: 420–30; Behmann 1934: 8–10). Thus, suppose the geometrical terms 'sphere' and 'includes' to be undefined on the basis of logical expressions, and suppose all further geometrical expressions defined on the basis of logical expressions together with 'sphere' and 'includes', as with Huntington (1913: 522–59). Let Huntington's postulates for (Euclidean) geometry, and all the theorems, be expanded by thoroughgoing replacement of definienda by definientia, so that they come to contain only logical expressions and 'sphere' and 'includes', and let the conjunction of the thus expanded postulates be represented as 'Hunt (sphere, includes).' Then, where 'Φ (sphere, includes)' is any of the theorems, similarly expanded into primitive terms, the point of view under consideration is that 'Φ (sphere, includes),' insofar as it is conceived as a mathematical truth, is to be construed as an ellipsis for 'If Hunt (spheres, includes) then Φ (sphere, includes).' Since 'Φ (sphere, includes)' is a logical consequence of Huntington's postulates, the above hypothetical statement is a truth of logic; it involves the expressions 'sphere' and 'includes' inessentially, in fact vacuously, since the logical deducibility of the theorems from the postulates is independent of the meanings of 'sphere' and 'includes' and survives the replacement of those expressions by any other grammatically admissible expressions whatever. Since, granted the fitness of Huntington's postulates, all and only those geometrical statements are truths of geometry which are logical consequences in this fashion of 'Hunt (sphere, includes),' all geometry becomes logic when interpreted in the above manner as a conventional ellipsis for a body of hypothetical statements.

But if, as a truth of mathematics, 'Φ (sphere, includes)' is short for 'If Hunt (sphere, includes) then Φ (sphere, includes),' still there remains, as part of this expanded statement, the original statement 'Φ (sphere, includes)'; this remains as a presumably true statement within some body of doctrine, say for the moment "non-mathematical geometry," even if the title of mathematical truth be restricted to the entire hypothetical statement in question. The body of all such hypothetical statements, describable as the "theory of deduction of non-mathematical geometry," is of course a part of logic; but the same is true of any "theory of deduction of sociology," "theory of deduction of Greek mythology," etc., which we might construct in parallel fashion with the aid of any set of postulates suited to sociology or to Greek mythology. The point of view toward geometry which is under consideration thus reduces merely to an exclusion of geometry from mathematics, a relegation of geometry to the status of sociology or Greek mythology; the labelling of the "theory of deduction of non-mathematical geometry" as "mathematical geometry" is a verbal *tour de force* which is equally applicable in the case of sociology or Greek mythology. To incorporate mathematics into logic by regarding all recalcitrant mathematical truths as elliptical hypothetical statements is thus in effect merely to restrict the term 'mathematics' to exclude those recalcitrant branches. But we are not interested in renaming. Those disciplines, geometry and the rest, which have traditionally been grouped under mathematics are the objects of the present discussion, and it is with the doctrine that mathematics in this sense is logic that we are here concerned.[3]

Discarding this alternative and returning, then, we see that if some mathematical expressions resist definition on the basis of logical ones, mathematics will reduce to logic only if, under a literal reading and without the gratuitous annexation of hypotheses, every mathematical truth contains (or is an abbreviation of one which contains) such recalcitrant expressions only inessentially if at all. But a mathematical expression

[3]Obviously the foregoing discussion has no bearing upon postulate method as such, nor upon Huntington's work.

sufficiently troublesome to have resisted trivial contextual definition in terms of logic can hardly be expected to occur thus idly in all its mathematical contexts. It would thus appear that for the tenability of the thesis that mathematics is logic it is not only sufficient but also necessary that all mathematical expressions be capable of definition on the basis solely of logical ones.

Though in framing logical definitions of mathematical expressions the ultimate objective be to make all mathematical truths logical truths, attention is not to be confined to mathematical and logical truths in testing the conformity of the definitions to usage. Mathematical expressions belong to the general language, and they are to be so defined that all statements containing them, whether mathematical truths, historical truths, or falsehoods under traditional usage, come to be construed as abbreviations of other statements which are correspondingly true or false. The definition introducing 'plus' must be such that the mathematical truth 'Smith's age plus Brown's equals Brown's age plus Smith's' becomes an abbreviation of a logical truth, as observed earlier; but it must also be such that 'Smith's age plus Brown's age equals Jones' age' becomes an abbreviation of a statement which is empirically true or false in conformity with the county records and the traditional usage of 'plus'. A definition which fails in this latter respect is no less Pickwickian than one which fails in the former; in either case nothing is achieved beyond the transient pleasure of a verbal recreation.

But for these considerations, contextual definitions of any mathematical expressions whatever could be framed immediately in purely logical terms, on the basis of any set of postulates adequate to the branch of mathematics in question. Thus, consider again Huntington's systematization of geometry. It was remarked that, granted the fitness of Huntington's postulates, a statement will be a truth of geometry if and only if it is logically deducible from 'Hunt (sphere, includes)' without regard to the meanings of 'sphere' and 'includes'. Thus 'Φ (sphere, includes)' will be a truth of geometry if and only if the following is a truth of logic: 'If α is any class and R any relation such that Hunt (α, R), then Φ (α, R).' For 'sphere' and 'includes' we might then adopt the following contextual definition: Where '———' is any statement containing 'α' or 'R' or both, let the statement 'If α is any class and R any relation such that Hunt (α, R), then ———' be abbreviated as that expression which is got from '———' by putting 'sphere' for 'α' and 'includes' for 'R' throughout. (In the case of a compound statement involving 'sphere' and 'includes', this definition does not specify whether it is the entire statement or each of its constituent statements that is to be accounted as shorthand in the described fashion; but this ambiguity can be eliminated by stipulating that the convention apply only to whole contexts.) 'Sphere' and 'includes' thus receive contextual definition in terms exclusively of logic, for any statement containing one or both of those expressions is construed by the definition as an abbreviation of a statement containing only logical expressions (plus whatever expressions the original statement may have contained other than 'sphere' and 'includes'). The definition satisfies past usage of 'sphere' and 'includes' to the extent of verifying all truths and falsifying all falsehoods of geometry; all those statements of geometry which are true, and only those, become abbreviations of truths of logic.

The same procedure could be followed in any other branch of mathematics, with the help of a satisfactory set of postulates for the branch. Thus nothing further would appear to be wanting for the thesis that mathematics is logic. And the royal road runs beyond that thesis, for the described method of logicizing a mathematical discipline can be applied likewise to any nonmathematical theory. But the whole procedure rests on failure to conform the definitions to usage; what is logicized is not the intended subject-matter. It is readily seen, e.g., that the suggested contextual definition of 'sphere' and 'includes', though transforming purely geometrical truths and falsehoods respectively into logical truths and falsehoods, transforms certain empirical truths into falsehoods and vice versa. Consider, e.g., the true statement 'A baseball is roughly a sphere,' more rigorously 'The whole of a baseball, except for a certain very thin, irregular peripheral layer, constitutes a sphere.' According to the contextual definition, this statement is an abbreviation for the following: 'If α is any class and R any relation such that Hunt (α, R), then the whole of a baseball, except for a thin peripheral layer, constitutes an [a member of] α.' This tells us that the whole of a baseball, except for a thin peripheral layer, belongs to every class α for which a relation R can be found such that Huntington's postulates are true of α and R. Now it happens that 'Hunt $(\alpha$, includes$)$' is true not only when α is taken as the class of all spheres, but also when α is restricted to the class of spheres a foot or more in diameter (cf. Huntington 1913: 540); yet the whole of a baseball, except for a thin peripheral layer, can hardly be said to constitute a sphere a foot or more in diameter. The statement is therefore false, whereas the preceding statement, supposedly an abbreviation of this one, was true under ordinary usage of words. The thus logicized rendering of any other discipline can be shown in analogous fashion to yield the sort of discrepancy observed just now for geometry, provided only that the postulates of the discipline admit, like those of geometry, of

alternative applications; and such multiple applicability is to be expected of any postulate set.[4]

Definition of mathematical notions on the basis of logical ones is thus a more arduous undertaking than would appear from a consideration solely of the truths and falsehoods of pure mathematics. Viewed *in vacuo*, mathematics is trivially reducible to logic through erection of postulate systems into contextual definitions; but "cette science n'a pas uniquement pour objet de contempler éternellement son propre nombril" (Poincaré 1908b: 199). When mathematics is recognized as capable of use, and as forming an integral part of general language, the definition of mathematical notions in terms of logic becomes a task whose completion, if theoretically possible at all, calls for mathematical genius of a high order. It was primarily to this task that Whitehead and Russell addressed themselves in their *Principia Mathematica*. They adopt a meager logical language as primitive, and on its basis alone they undertake to endow mathematical expressions with definitions which conform to usage in the full sense described above: definitions which not only reduce mathematical truths and falsehoods to logical ones, but reduce *all* statements, containing the mathematical expressions in question, to equivalent statements involving logical expressions instead of the mathematical ones. Within *Principia* the program has been advanced to such a point as to suggest that no fundamental difficulties stand in the way of completing the process. The foundations of arithmetic are developed in *Principia*, and therewith those branches of mathematics are accommodated which, like analysis and theory of number, spring from arithmetic. Abstract algebra proceeds readily from the relation theory of *Principia*. Only geometry remains untouched, and this field can be brought into line simply by identifying n-dimensional figures with those n-adic arithmetical relations ("equations in n variables") with which they are correlated through analytic geometry (cf. Study 1914: 86–92). Some question Whitehead and Russell's reduction of mathematics to logic (cf., e.g., Dubislav 1925: 193–208; Hilbert 1928: 12, 21), on grounds for whose exposition and criticism there is not space; the thesis that all mathematics reduces to logic is, however, substantiated by *Principia* to a degree satisfactory to most of us. There is no need here to adopt a final stand in the matter.

If for the moment we grant that all mathematics is thus definitionally constructible from logic, then mathematics becomes true by convention

[4]Note that a postulate set is superfluous if it *demonstrably* admits of one and only one application: for it then embodies an adequate defining property for each of its constituent primitive terms. Cf. Tarski 1935a: 85 (proposition 2).

in a relative sense: mathematical truths become conventional transcriptions of logical truths. Perhaps this is all that many of us mean to assert when we assert that mathematics is true by convention; at least, an *analytic* statement is commonly explained merely as one which proceeds from logic and definitions, or as one which, on replacement of definienda by definientia, becomes a truth of logic.[5] But in strictness we cannot regard mathematics as true purely by convention unless all those logical principles to which mathematics is supposed to reduce are likewise true by convention. And the doctrine that mathematics is *analytic* accomplishes a less fundamental simplification for philosophy than would at first appear, if it asserts only that mathematics is a conventional transcription of logic and not that logic is convention in turn: for if in the end we are to countenance any *a priori* principles at all which are independent of convention, we should not scruple to admit a few more, nor attribute crucial importance to conventions which serve only to diminish the number of such principles by reducing some to others.

But if we are to construe logic also as true by convention, we must rest logic ultimately upon some manner of convention other than definition: for it was noted earlier that definitions are available only for transforming truths, not for founding them. The same applies to any truths of mathematics which, contrary to the supposition of a moment ago, may resist definitional reduction to logic; if such truths are to proceed from convention, without merely being reduced to antecedent truths, they must proceed from conventions other than definitions. Such a second sort of convention, generating truths rather than merely transforming them, has long been recognized in the use of postulates.[6] Application of this method to logic will occupy the next section; customary ways of rendering postulates and rules of inference will be departed from, however, in favor of giving the whole scheme the explicit form of linguistic convention.

II

Let us suppose an approximate maximum of definition to have been accomplished for logic, so that we are left with about as meager as possible an array of primitive notational devices. There are indefinitely many ways of framing the definitions, all conforming to the same usage of the expressions in question; apart from the objective of defining much

[5] Cf. Frege 1884: 4; Behmann 1934: 5. Carnap, 1934b, uses the term in essentially the same sense but subject to more subtle and rigorous treatment.

[6] The function of postulates as conventions seems to have been first recognized by Gergonne (1819). His designation of them as 'implicit definitions,'' which has had some following in the literature, is avoided here.

in terms of little, choice among these ways is guided by convenience or chance. Different choices involve different sets of primitives. Let us suppose our procedure to be such as to reckon among the primitive devices the *not*-idiom, the *if*-idiom ('If ... then ...'), the *every*-idiom ('No matter what x may be, ———x———'), and one or two more as required. On the basis of this much, then, all further logical notation is to be supposed defined; all statements involving any further logical notation become construed as abbreviations of statements whose logical constituents are limited to those primitives.

'Or', as a connective joining statements to form new statements, is amenable to the following contextual definition in terms of the *not*-idiom and the *if*-idiom: A pair of statements with 'or' between is an abbreviation of the statement made up successively of these ingredients: first, 'If'; second, the first statement of the pair, with 'not' inserted to govern the main verb (or, with 'it is false that' prefixed); third, 'then'; fourth, the second statement of the pair. The convention becomes clearer if we use the prefix '$\sim$' as an artificial notation for denial, thus writing '$\sim$ice is hot' instead of 'Ice is not hot' or 'It is false that ice is hot.' Where '———' and '——' are any statements, our definition then introduces '——— or ——' as an abbreviation of 'If $\sim$——— then ——.' Again 'and', as a connective joining statements, can be defined contextually by construing '——— and ——' as an abbreviation for '$\sim$ if ——— then $\sim$——.' Every such idiom is what is known as a *truth-function*, and is characterized by the fact that the truth or falsehood of the complex statement which it generates is uniquely determined by the truth or falsehood of the several statements which it combines. All truth-functions are known to be constructible in terms of the *not*- and *if*-idioms as in the above examples.[7] On the basis of the truth-functions, then, together with our further primitives – the *every*-idiom *et al.* – all further logical devices are supposed defined.

A word may, through historical or other accidents, evoke a train of ideas bearing no relevance to the truth or falsehood of its context; in point of *meaning*, however, as distinct from connotation, a word may be said to be determined to whatever extent the truth or falsehood of its contexts is determined. Such determination of truth or falsehood may be outright, and to that extent the meaning of the word is absolutely determined; or it may be relative to the truth or falsehood of statements containing other words, and to that extent the meaning of the word is determined relatively to those other words. A definition endows a word with

[7] Sheffer (1913: 481–8) has shown ways of constructing these two, in turn, in terms of one; strictly, therefore, such a one should supplant the two in our ostensibly minimal set of logical primitives. Exposition will be facilitated, however, by retaining the redundancy.

complete determinacy of meaning relative to other words. But the alternative is open to us, on introducing a new word, of determining its meaning *absolutely* to whatever extent we like by specifying contexts which are to be true and contexts which are to be false. In fact, we need specify only the former: for falsehood may be regarded as a derivative property depending on the word '~', in such wise that falsehood of '———' means simply truth of '~———.' Since all contexts of our new word are meaningless to begin with, neither true nor false, we are free to run through the list of such contexts and pick out as true such ones as we like; those selected become true by fiat, by linguistic convention. For those who would question them we have always the same answer, 'You use the word differently.' The reader may protest that our arbitrary selection of contexts as true is subject to restrictions imposed by the requirement of *consistency* – e.g., that we must not select both '———' and '~———'; but this consideration, which will receive a clearer status a few pages hence, will be passed over for the moment.

Now suppose in particular that we abstract from existing usage of the locutions 'if-then', 'not' (or '~'), and the rest of our logical primitives, so that for the time being these become meaningless marks, and the erstwhile statements containing them lose their status as statements and become likewise meaningless, neither true nor false; and suppose we run through all those erstwhile statements, or as many of them as we like, segregating various of them arbitrarily as true. To whatever extent we carry this process, we to that extent determine meaning for the initially meaningless marks 'if', 'then', '~', and the rest. Such contexts as we render true are true by convention.

We saw earlier that if all expressions occurring essentially in a true statement S and not belonging to a class α are given definitions in terms solely of members of α, then S becomes a definitional abbreviation of a truth S'' involving only members of a α essentially. Now let α comprise just our logical primitives, and let S be a statement which, under ordinary usage, is true and involves only logical expressions essentially. Since all logical expressions other than the primitives are defined in terms of the primitives, it then follows that S is an abbreviation of a truth S'' involving only the primitives essentially. But if one statement S is a definitional abbreviation of another S'', the truth of S proceeds wholly from linguistic convention if the truth of S'' does so. Hence if, in the above process of arbitrarily segregating statements as true by way of endowing our logical primitives with meaning, *we assign truth to those statements which, according to ordinary usage, are true and involve only our primitives essentially*, then not only will the latter statements be true by convention, but so will all statements which are true under ordinary usage and involve only logical expressions essentially. Since, as remarked earlier, every logical truth involves (or is an abbreviation of another which involves) only logical expressions essentially, the described scheme of assigning truth makes all logic true by convention.

Not only does such assignment of truth suffice to make all those statements true by convention which are true under ordinary usage and involve only logical expressions essentially, but it serves also to make all those statements false by convention which are false under ordinary usage and involve only logical expressions essentially. This follows from our explanation of the falsehood of '———' as the truth of '~———', since '———' will be false under ordinary usage if and only if '~———' is true under ordinary usage. The described assignment of truth thus goes far towards fixing all logical expressions in point of meaning, and fixing them in conformity with usage. Still many statements containing logical expressions remain unaffected by the described assignments: all those statements which, from the standpoint of ordinary usage, involve some non-logical expressions essentially. There is hence room for supplementary conventions of one sort or another over and above the described truth-assignments, by way of completely fixing the meanings of our primitives – and fixing them, it is to be hoped, in conformity with ordinary usage. Such supplementation need not concern us now; the described truth-assignments provide partial determinations which, as far as they go, conform to usage and which go far enough to make all logic true by convention.

But we must not be deceived by schematism. It would appear that we sit down to a list of expressions and check off as arbitrarily true all those which, under ordinary usage, are true statements involving only our logical primitives essentially; but this picture wanes when we reflect that the number of such statements is infinite. If the convention whereby those statements are singled out as true is to be formulated in finite terms, we must avail ourselves of conditions finite in length which determine infinite classes of expressions.[8]

Such conditions are ready at hand. One, determining an infinite class of expressions all of which, under ordinary usage, are true statements involving only our primitive *if*-idiom essentially, is the condition of being obtainable from

[8]Such a condition, if effective, constitutes a *formal system*. Usually we assign such meanings to the signs as to construe the expressions of the class as statements, specifically true statements, theorems; but this is neither intrinsic to the system nor necessary in all cases for a useful application of the system.

(1) 'If if p then q then if if q then r then if p then r'

by putting a statement for 'p', a statement for 'q', and a statement for 'r'. In more customary language the form (1) would be expanded, for clarity, in some such fashion as this: 'If it is the case that if p then q, then, if it is the case further that if q then r, then, if p, r.' The form (1) is thus seen to be the principle of the syllogism. Obviously it is true under ordinary usage for all substitutions of statements for 'p', 'q', and 'r'; hence such results of substitution are, under ordinary usage, true statements involving only the *if*-idiom essentially. One infinite part of our program of assigning truth to all expressions which, under ordinary usage, are true statements involving only our logical primitives essentially, is thus accomplished by the following convention:

(I) Let all results of putting a statement for 'p', a statement for 'q', and a statement for 'r' in (1) be true.

Another infinite part of the program is disposed of by adding this convention:

(II) Let any expression be true which yields a truth when put for 'q' in the result of putting a truth for 'p' in 'If p then q.'

Given truths '———' and 'If ——— then ——', (II) yields the truth of '——'. That (II) conforms to usage, i.e., that from statements which are true under ordinary usage (II) leads only to statements which are likewise true under ordinary usage, is seen from the fact that under ordinary usage a statement '——' is always true if statements '———' and 'If ——— then ——' are true. Given all the truths yielded by (I), (II) yields another infinity of truths which, like the former, are under ordinary usage truths involving only the *if*-idiom essentially. How this comes about is seen roughly as follows. The truths yielded by (I), being of the form of (1), are complex statements of the form 'If ——— then ——'. The statement '———' here may in particular be of the form (1) in turn, and hence likewise be true according to (I). Then, by (II), '——' becomes true. In general '——' will not be of the form (1), hence would not have been obtainable by (I) alone. Still '——' will in every such case be a statement which, under ordinary usage, is true and involves only the *if*-idiom essentially; this follows from the observed conformity of (I) and (II) to usage, together with the fact that the above derivation of '——' demands nothing of '——' beyond proper structure in terms of 'if-then'. Now our stock of truths embraces not only those yielded by (I) alone, i.e., those having the form (1), but also all those thence derivable by (II) in the

manner in which '——' has just now been supposed derived.[9] From this increased stock we can derive yet further ones by (II), and these likewise will, under ordinary usage, be true and involve only the *if*-idiom essentially. The generation proceeds in this fashion *ad infinitum*.

When provided only with (I) as an auxiliary source of truth, (II) thus yields only truths which under ordinary usage are truths involving only the *if*-idiom essentially. When provided with further auxiliary sources of truths, however, e.g., the convention (III) which is to follow, (II) yields truths involving further locutions essentially. Indeed, the effect of (II) is not even confined to statements which, under ordinary usage, involve only logical locutions essentially; (II) also legislates regarding other statements, to the extent of specifying that no two statements '———' and 'If ——— then ——' can both be true unless '——' is true. But this overflow need not disturb us, since it also conforms to ordinary usage. In fact, it was remarked earlier that room remained for supplementary conventions, over and above the described truth-assignments, by way of further determining the meanings of our primitives. This overflow accomplishes just that for the *if*-idiom; it provides, with regard even to a statement 'If ——— then ——' which from the standpoint of ordinary usage involves non-logical expressions essentially, that the statement is not to be true if '———' is true and '——' not.

But present concern is with statements which, under ordinary usage, involve only our logical primitives essentially; by (I) and (II) we have provided for the truth of an infinite number of such statements, but by no means all. The following convention provides for the truth of another infinite set of such statements; these, in contrast to the preceding, involve not only the *if*-idiom but also the *not*-idiom essentially (under ordinary usage).

(III) Let all results of putting a statement for 'p' and a statement for 'q', in 'If p then if $\sim p$ then q' or 'If if $\sim p$ then p then p,' be true.[10]

Statements generated thus by substitution in 'If p then if $\sim p$ then q' are statements of hypothetical form in which two mutually contradictory statements occur as premisses; obviously such statements are trivially true, under ordinary usage, no matter what may figure as conclusion. Statements generated by substitution in 'If [it is the case that] if $\sim p$ then p, then p' are likewise true under ordinary usage, for one reasons as

[9] The latter in fact comprise all and only those statements which have the form 'If if if if q then r then if p then r then s then if if p then q then s'.

[10] (1) and the two formulae in (III) are Łukasiewicz's three postulates for the propositional calculus.

follows: Grant the hypothesis, viz., that if $\sim p$ then p; then we must admit the conclusion, viz., that p, since even denying it we admit it. Thus all the results of substitution referred to in (III) are true under ordinary usage no matter what the substituted statements may be; hence such results of substitution are, under ordinary usage, true statements, involving nothing essentially beyond the *if*-idiom and the *not*-idiom ('$\sim$').

From the infinity of truths adopted in (III), together with those already at hand from (I) and (II), infinitely more truths are generated by (II). It happens, curiously enough, that (III) adds even to our stock of statements which involve only the *if*-idiom essentially (under ordinary usage); there are truths of that description which, though lacking the *not*-idiom, are reached by (I)–(III) and not by (I) and (II). This is true, e.g., of any instance of the principle of identity, say

(2) 'If time is money then time is money'.

It will be instructive to derive (2) from (I)–(III), as an illustration of the general manner in which truths are generated by those conventions. (III), to begin with, directs that we adopt these statements as true:

(3) 'If time is money then if time is not money then time is money'.
(4) 'If if time is not money then time is money then time is money'.

(I) directs that we adopt this as true:

(5) 'If if time is money then if time is not money then time is money then if if if time is not money then time is money then time is money then if time is money then time is money'.

(II) tells us that, in view of the truth of (5) and (3), this is true:

(6) 'If if if time is not money then time is money then time is money then if time is money then time is money'.

Finally (II) tells us that, in view of the truth of (6) and (4), (2) is true.

If a statement S is generated by (I)–(III), obviously only the structure of S in terms of 'if-then' and '$\sim$' was relevant to the generation; hence all those variants S_i of S which are obtainable by any grammatically admissible substitutions upon constituents of S not containing 'if', 'then', or '$\sim$', are likewise generated by (I)–(III). Now it has been observed that (I)–(III) conform to usage, i.e., generate only statements which are true under ordinary usage; hence S and all the S_i are uniformly true under ordinary usage, the S_i are therefore vacuous variants of S, and hence only 'if', 'then', and '$\sim$' occur essentially in S. Thus (I)–(III)

generate only statements which under ordinary usage are truths involving only the *if*-idiom and the *not*-idiom essentially.

It can be shown also that (I)–(III) generate *all* such statements.[11] Consequently (I)–(III), aided by our definitions of logical locutions in terms of our primitives, are adequate to the generation of all statements which under ordinary usage are truths which involve any of the so-called truth-functions but nothing else essentially: for it has been remarked that all the truth-functions are definable on the basis of the *if*-idiom and the *not*-idiom. All such truths thus become true by convention. They comprise all those statements which are instances of any of the principles of the so-called propositional calculus.

To (I)–(III) we may now add a further convention or two to cover another of our logical primitives – say the *every*-idiom. A little more in this direction, by way of providing for our remaining primitives, and the program is completed; all logic, in some sense, becomes true by convention. The conventions with which (I)–(III) would thus be supplemented would be more complex than (I)–(III). The set of conventions would be an adaptation of one of the various existing systematizations of general logistic, in the same way in which (I)–(III) are an adaptation of a systematization of the propositional calculus.

The systematization chosen must indeed leave some logical statements undecided, by Gödel's theorem, if we set generous bounds to the logical vocabulary. But no matter; logic still becomes true by convention insofar as it gets reckoned as true on any account.

Let us now consider the protest which the reader raised earlier, viz., that our freedom in assigning truth by convention is subject to restrictions imposed by the requirement of consistency (see, e.g., Poincaré 1908b: 162–3, 195–8; Schlick 1918, 2nd ed.: 36, 327). Under the fiction, implicit in an earlier stage of our discussion, that we check off our truths one by one in an exhaustive list of expressions, consistency in the assign-

[11] The proof rests essentially upon Łukasiewicz's proof (1929) that his three postulates for the propositional calculus, viz., (1) and the formulae in (III), are *complete*. Adaptation of his result to present purposes depends upon the fact, readily established, that any formula generable by his two rules of inference (the so-called rule of substitution and a rule answering to (II)) can be generated by applying the rules in such order that all applications of the rule of substitution precede all applications of the other rule. This fact is relevant because of the manner in which the rule of substitution has been absorbed, here, into (I) and (III). The adaptation involves also two further steps, which however present no difficulty: we must make connection between Łukasiewicz's *formulae*, containing variables 'p', 'q', etc., and the concrete *statements* which constitute the present subject-matter; also between *completeness*, in the sense (Post's) in which Łukasiewicz uses the term, and the generability of all statements which under ordinary usage are truths involving only the *if*-idiom or the *not*-idiom essentially.

ment of truth is nothing more than a special case of conformity to usage. If we make a mark in the margin opposite an expression '———', and another opposite '~———', we sin only against the established usage of '~' as a denial sign. Under the latter usage '———' and '~———' are not both true; in taking them both by convention as true we merely endow the sign '~', roughly speaking, with a meaning other than denial. Indeed, we might so conduct our assignments of truth as to allow no sign of our language to behave analogously to the denial locution of ordinary usage; perhaps the resulting language would be inconvenient, but conventions are often inconvenient. It is only the objective of ending up with our mother tongue that dissuades us from marking both '———' and '~———', and this objective would dissuade us also from marking 'It is always cold on Thursday.'

The requirement of consistency still retains the above status when we assign truth wholesale through general conventions such as (I)–(III). Each such convention assigns truth to an infinite sheaf of the entries in our fictive list, and in this function the conventions cannot conflict; by overlapping in their effects they reinforce one another, by not overlapping they remain indifferent to one another. If some of the conventions specified entries to which truth was *not* to be assigned, genuine conflict might be apprehended; such negative conventions, however, have not been suggested. (II) was, indeed, described earlier as specifying that 'If ——— then ———' is not to be true if '———' is true and '———' not; but within the framework of the conventions of truth-assignment this apparent proscription is ineffectual without antecedent proscription of '———'. Thus any inconsistency among the general conventions will be of the sort previously considered, viz., the arbitrary adoption of both '———' and '~———' as true; and the adoption of these was seen merely to impose some meaning other than denial upon the sign '~'. As theoretical restrictions upon our freedom in the conventional assignment of truth, requirements of consistency thus disappear. Preconceived usage may lead us to stack the cards, but does not enter the rules of the game.

III

Circumscription of our logical primitives in point of meaning, through conventional assignment of truth to various of their contexts, has been seen to render all logic true by convention. Then if we grant the thesis that mathematics is logic, i.e., that all mathematical truths are definitional abbreviations of logical truths, it follows that mathematics is true by convention.

If on the other hand, contrary to the thesis that mathematics is logic, some mathematical expressions resist definition in terms of logical ones, we can extend the foregoing method into the domain of these recalcitrant expressions: we can circumscribe the latter through conventional assignment of truth to various of their contexts, and thus render mathematics conventionally true in the same fashion in which logic has been rendered so. Thus, suppose some mathematical expressions to resist logical definition, and suppose them to be reduced to as meager as possible a set of mathematical primitives. In terms of these and our logical primitives, then, all further mathematical devices are supposed defined; all statements containing the latter become abbreviations of statements containing by way of mathematical notation only the primitives. Here, as remarked earlier in the case of logic, there are alternative courses of definition and therewith alternative sets of primitives; but suppose our procedure to be such as to count 'sphere' and 'includes' among the mathematical primitives. So far we have a set of conventions, (I)–(III) and a few more, let us call them (IV)–(VII), which together circumscribe our logical primitives and yield all logic. By way of circumscribing the further primitives 'sphere' and 'includes', let us now add this convention to the set:

(VIII) Let 'Hunt (sphere, includes)' be true.

Now we saw earlier that where 'Φ (sphere, includes)' is any truth of geometry, supposed expanded into primitive terms, the statement

(7) 'If Hunt (sphere, includes) then Φ (sphere, includes);

is a truth of logic. Hence (7) is one of the expressions to which truth is assigned by the conventions (I)–(VII). Now (II) instructs us, in view of convention (VIII) and the truth of (7), to adopt 'Φ (sphere, includes)' as true. In this way each truth of geometry is seen to be present among the statements to which truth is assigned by the conventions (I)–(VII).

We have considered four ways of construing geometry. One way consisted of straightforward definition of geometrical expressions in terms of logical ones, within the direction of development represented by *Principia Mathematica*; this way, presumably, would depend upon identification of geometry with algebra through the correlations of analytic geometry, and definition of algebraic expressions on the basis of logical ones as in *Principia Mathematica*. By way of concession to those who have fault to find with certain technical points in *Principia*, this possibility was allowed to retain a tentative status. The other three ways all made use of Huntington's postulates, but are sharply to be distinguished from one another. The first was to include geometry in logic by construing geometrical truths as elliptical for hypothetical statements bearing 'Hunt (sphere, includes)' as hypothesis; this was seen to be a mere evasion, tantamount, under its verbal disguise, to the concession that geometry is not logic after all. The next procedure was to define 'sphere' and 'includes' contextually in terms of logical expressions by construing 'Φ (sphere, includes)' in every case as an abbreviation of 'If α is any class and R any relation such that Hunt (α, R), then Φ (α, R)'. This definition was condemned on the grounds that it fails to yield the intended usage of the defined terms. The last procedure finally, just now presented, renders geometry true by convention without making it part of logic. Here 'Hunt (sphere, includes)' is made true by fiat, by way of conventionally delimiting the meanings of "sphere" and "includes." The truths of geometry then emerge not as truths of logic, but in parallel fashion to the truths of logic.

This last method of accommodating geometry is available also for any other branch of mathematics which may resist definitional reduction to logic. In each case we merely set up a conjunction of postulates for that branch as true by fiat, as a conventional circumscription of the meanings of the constituent primitives, and all the theorems of the branch thereby become true by convention: the convention thus newly adopted together with the conventions (I)–(VII). In this way all mathematics becomes conventionally true, not by becoming a definitional transcription of logic, but by proceeding from linguistic convention in the same way as does logic.

But the method can even be carried beyond mathematics, into the so-called empirical sciences. Having framed a maximum of definitions in the latter realm, we can circumscribe as many of our "empirical" primitives as we like by adding further conventions to the set adopted for logic and mathematics; a correspondingly portion of "empirical" science then becomes conventionally true in precisely the manner observed above for geometry.

The impossibility of defining any of the "empirical" expressions in terms exclusively of logical and mathematical ones may be recognized at the outset: for if any proved to be so definable, there can be no question but that it would thenceforward be recognized as belonging to pure mathematics. On the other hand vast numbers of "empirical" expressions are of course definable on the basis of logical and mathematical ones together with other "empirical" ones. Thus 'momentum' is defined as 'mass times velocity'; 'event' may be defined as 'referent of the *later*-relation', i.e., 'whatever is later than something'; 'instant' may be defined as 'maximal class of events no one of which is later than any

other event of the class' (Russell 1914: 126); 'time' as 'the class of all instants'; and so on. In these examples 'momentum' is defined on the basis of mathematical expressions together with the further expressions 'mass' and 'velocity'; 'event', 'instant', and 'time' are all defined on the basis ultimately of logical expressions together with the one further expression 'later than'.

Now suppose definition to have been performed to the utmost among such non-logical, non-mathematical expressions, so that the latter are reduced to as few "empirical" primitives as possible.[12] *All* statements then become abbreviations of statements containing nothing beyond the logical and mathematical primitives and these "empirical" ones. Here, as before, there are alternatives of definition and therewith alternative sets of primitives; but supppose our primitives to be such as to include 'later than', and consider the totality of those statements which under ordinary usage are truths involving only 'later than' and mathematical or logical expressions essentially. Examples of such statements are 'Nothing is later than itself'; 'If Pompey died later than Brutus and Brutus died later than Caesar then Pompey died later than Caesar.' All such statements will be either very general principles, like the first example, or else instances of such principles, like the second example. Now it is a simple matter to frame a small set of general statements from which all and only the statements under consideration can be derived by means of logic and mathematics. The conjunction of these few general statements can then be adopted as true by fiat, as 'Hunt (sphere, includes)' was adopted in (VIII); their adoption is a conventional circumscription of the meaning of the primitive 'later than'. Adoption of this convention renders all those statements conventionally true which under ordinary usage are truths essentially involving any logical or mathematical expressions, or 'later than', or any of the expressions which, like 'event', 'instant', and 'time', are defined on the basis of the foregoing, and inessentially involving anything else.

Now we can pick another of our "empirical" primitives, perhaps 'body' or 'mass' or 'energy', and repeat the process. We can continue in this fashion to any desired point, circumscribing one primitive after another by convention, and rendering conventionally true all statements which under ordinary usage are truths essentially involving only the locutions treated up to that point. If in disposing successively of our

[12] Carnap has pursued his program with such amazing success as to provide grounds for expecting all the expressions to be definable ultimately in terms of logic and mathematics plus just one "empirical" primitive, representing a certain dyadic relation described as *recollection of resemblance* (1928a). But for the present cursory considerations no such spectacular reducibility need be presupposed.

"empirical" primitives in the above fashion we take them up in an order roughly describable as leading from the general to the special, then as we progress we may expect to have to deal more and more with statements which are true under ordinary usage only with reservations, only with a probability recognized as short of certainty. But such reservations need not deter us from rendering a statement true by convention; so long as under ordinary usage the presumption is rather for than against the statement, our convention conforms to usage in verifying it. In thus elevating the statement from putative to conventional truth, we still retain the right to falsify the statement tomorrow if those events should be observed which would have occasioned its repudiation while it was still putative: for conventions are commonly revised when new observations show the revision to be convenient.

If in describing logic and mathematics as true by convention what is meant is that the primitives *can* be conventionally circumscribed in such fashion as to generate all and only the so-called truths of logic and mathematics, the characterization is empty; our last considerations show that the same might be said of any other body of doctrine as well. If on the other hand it is meant merely that the speaker adopts such conventions for those fields but not for others, the characterization is uninteresting; while if it is meant that it is a general practice to adopt such conventions explicitly for those fields but not for others, the first part of the characterization is false.

Still, there is the apparent contrast between logico-mathematical truths and others that the former are *a priori*, the latter *a posteriori*; the former have "the character of an inward necessity," in Kant's phrase, the latter do not. Viewed behavioristically and without reference to a metaphysical system, this contrast retains reality as a contrast between more and less firmly accepted statements; and it obtains antecedently to any *post facto* fashioning of conventions. There are statements which we choose to surrender last, if at all, in the course of revamping our sciences in the face of new discoveries; and among these there are some which we will not surrender at all, so basic are they to our whole conceptual scheme. Among the latter are to be counted the so-called truths of logic and mathematics, regardless of what further we may have to say of their status in the course of a subsequent sophicated philosophy. Now since these statements are destined to be maintained independently of our observations of the world, we may as well make use here of our technique of conventional truth-assignment and thereby forestall awkward metaphysical questions as to our *a priori* insight into necessary truths. On the other hand this purpose would not motivate extension of the truth-assignment process into the realm of erstwhile contingent state-

ments. On such grounds, then, logic and mathematics may be held to be conventional while other fields are not; it may be held that it is philosophically important to circumscribe the logical and mathematical primitives by conventions of truth-assignment which yield logical and mathematical truths, but that it is idle elaboration to carry the process further. Such a characterization of logic and mathematics is perhaps neither empty nor uninteresting nor false.

In the adoption of the very conventions (I)–(III) etc. whereby logic itself is set up, however, a difficulty remains to be faced. Each of these conventions is general, announcing the truth of every one of an infinity of statements conforming to a certain description; derivation of the truth of any specific statement from the general convention thus requires a logical inference, and this involves us in an infinite regress. E.g., in deriving (6) from (3) and (5) on the authority of (II) we *infer*, from the general announcement (II) and the specific premiss that (3) and (5) are true statements, the conclusion that

(7) (6) is to be true.

An examination of this inference will reveal the regress. For present purposes it will be simpler to rewrite (II) thus:

(II′) No matter what x may be, no matter what y may be, no matter what z may be, if x and z are true [statements] and z is the result of putting x for 'p' and y for 'q' in 'If p then q' then y is to be true.

We are to take (II′) as a premiss, then, and in addition the premiss that (3) and (5) are true. We may also grant it as known that (5) is the result of putting (3) for 'p' and (6) for 'q' in 'If p then q.' Our second premiss may thus be rendered compositely as follows:

(8) (3) and (5) are true and (5) is the result of putting (3) for 'p' and (6) for 'q' in 'If p then q.'

From these two premisses we propose to infer (7). This inference is obviously sound logic; as logic, however, it involves use of (II′) and others of the conventions from which logic is supposed to spring. Let us try to perform the inference on the basis of those conventions. Suppose that our convention (IV), passed over earlier, is such as to enable us to infer specific instances from statements which, like (II′), involve the *every*-idiom; i.e., suppose that (IV) entitles us in general to drop the prefix 'No matter what x [or y, etc.] may be' and simultaneously to introduce a concrete designation instead of 'x' [or 'y', etc.] in the sequel. By invoking (IV) three times, then, we can infer the following from (II′):

(9) If (3) and (5) are true and (5) is the result of putting (3) for '*p*' and (6) for '*q*' in 'If *p* then *q*' then (6) is to be true.

It remains to infer (7) from (8) and (9). But this is an inference of the kind for which (II′) is needed; from the fact that

(10) (8) and (9) are true and (9) is the result of putting (8) for '*p*' and (7) for '*q*' in 'If *p* then *q*'.

we are to infer (7) with help of (II′). But the task of getting (7) from (10) and (II′) is exactly analogous to our original task of getting (6) from (8) and (II′); the regress is thus under way (cf. Dodgson 1895: 278–80). (Incidentally the derivation of (9) from (II′) by (IV), granted just now for the sake of argument, would encounter a similar obstacle; so also the various unanalyzed steps in the derivation of (8).)

In a word, the difficulty is that if logic is to proceed *mediately* from conventions, logic is needed for inferring logic from the conventions. Alternatively, the difficulty which appears thus as a self-presupposition of doctrine can be framed as turning upon a self-presupposition of primitives. It is supposed that the *if*-idiom, the *not*-idiom, the *every*-idiom, and so on, mean nothing to us initially, and that we adopt the conventions (I)–(VII) by way of circumscribing their meaning; and the difficulty is that communication of (I)–(VII) themselves depends upon free use of those very idioms which we are attempting to circumscribe, and can succeed only if we are already conversant with the idioms. This becomes clear as soon as (I)–(VII) are rephrased in rudimentary language, after the manner of (II′).[13] It is important to note that this difficulty besets only the method of wholesale truth-assignment, not that of definition. It is true, e.g., that the contextual definition of 'or' presented at the beginning of the second section was communicated with the help of logical and other expressions which cannot be expected to have been endowed with meaning at the stage where logical expressions are first being introduced. But a definition has the peculiarity of being theoretically dispensable; it introduces a scheme of abbreviation, and we are free, if we like, to

[13]Incidentally the conventions presuppose also some further locutions, e.g., 'true' ('a true statement'), 'the result of putting ... for ... in ...', and various nouns formed by displaying expressions in quotation marks. The linguistic presuppositions can of course be reduced to a minimum by careful rephrasing; (II′), e.g., can be improved to the following extent:

(II′) No matter what *x* may be, no matter what *y* may be, no matter what *z* may be, if *x* is true then if *z* is true then if *z* is the result of putting *x* for '*p*' in the result of putting *y* for '*q*' in 'If *p* then *q*' then *y* is true.

This involves just the *every*-idiom, the *if*-idiom, 'is', and the further locutions mentioned above.

forego the brevity which it affords until enough primitives have been endowed with meaning, through the method of truth-assignment or otherwise, to accommodate full exposition of the definitio On the other hand the conventions of truth-assignment cannot be thus withheld until preparations are complete, because they are needed in the preparations.

If the truth-assignments were made one by one, rather than an infinite number at a time, the above difficulty would disappear; truths of logic such as (2) would simply be asserted severally by fiat, and the problem of inferring them from more general conventions would not arise. This course was seen to be closed to us, however, by the infinitude of the truths of logic.

It may still be held that the conventions (I)–(VIII), etc., are *observed* from the start, and that logic and mathematics thereby become conventional. It may be held that we can adopt conventions through behavior, without first announcing them in words; and that we can return and formulate our conventions verbally afterward, if we choose, when a full language is at our disposal. It may be held that the verbal formulation of conventions is no more a prerequisite of the adoption of the conventions than the writing of a grammar is a prerequisite of speech; that explicit exposition of conventions is merely one of many important uses of a completed language. So conceived, the conventions no longer involve us in vicious regress. Inference from general conventions is no longer demanded initially, but remains to the subsequent sophisticated stage where we frame general statements of the conventions and show how various specific conventional truths, used all along, fit into the general conventions as thus formulated.

It must be conceded that this account accords well with what we actually do. We discourse without first phrasing the conventions; afterwards, in writings such as this, we formulate them to fit our behavior. On the other hand it is not clear wherein an adoption of the conventions, antecedently to their formulation, consists; such behavior is difficult to distinguish from that in which conventions are disregarded. When we first agree to understand 'Cambridge' as referring to Cambridge in England failing a suffix to the contrary, and then discourse accordingly, the rôle of linguistic convention is intelligible; but when a convention is incapable of being communicated until after its adoption, its rôle is not so clear. In dropping the attributes of deliberateness and explicitness from the notion of linguistic convention we risk depriving the latter of any explanatory force and reducing it to an idle label. We may wonder what one adds to the bare statement that the truths of logic and mathematics are *a priori*, or to the still barer behavioristic statement that they are firmly accepted, when he characterizes them as true by convention in such a sense.

The more restricted thesis discussed in the first section, viz., that mathematics is a conventional transcription of logic, is far from trivial; its demonstration is a highly technical undertaking and an important one, irrespectively of what its relevance may be to fundamental principles of philosophy. It is valuable to show the reducibility of any principle to another through definition of erstwhile primitives, for every such achievement reduces the number of our presuppositions and simplifies and integrates the structure of our theories. But as to the larger thesis that mathematics and logic proceed wholly from linguistic conventions, only further clarification can assure us that this asserts anything at all.

Carnap and logical truth

W. V. QUINE

I

Kant's question "How are synthetic judgments a priori possible?" precipitated the *Critique of Pure Reason*. Question and answer notwithstanding, Mill and others persisted in doubting that such judgments were possible at all. At length some of Kant's own clearest purported instances, drawn from arithmetic, were sweepingly disqualified (or so it seemed; but see §II) by Frege's reduction of arithmetic to logic. Attention was thus forced upon the less tendentious and indeed logically prior question, "How is logical certainty possible?" It was largely this latter question that precipitated the form of empiricism which we associate with between-war Vienna – a movement which began with Wittgenstein's *Tractatus* and reached its maturity in the work of Carnap.

Mill's position on the second question had been that logic and mathematics were based on empirical generalizations, despite their superficial appearance to the contrary. This doctrine may well have been felt to do less than justice to the palpable surface differences between the deductive sciences of logic and mathematics, on the one hand, and the empirical sciences ordinarily so-called on the other. Worse, the doctrine derogated from the certainty of logic and mathematics; but Mill may not have been one to be excessively disturbed by such a consequence. Perhaps classical mathematics did lie closer to experience then than now; at any rate the infinitistic reaches of set theory, which are so fraught with speculation and so remote from any possible experience, were unexplored in his day. And it is against just these latter-day mathematical extravagances that empiricists outside the Vienna Circle have since been known to inveigh

Written early in 1954 for P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap* (La Salle, Ill.: Open Court, 1963) at the request of the editor. It appeared in Italian translation in *Rivista di Filosofia*, 1957, and selected portions amounting to somewhat less than half appeared also in Sidney Hook, ed., *American Philosophers at Work* (New York: Criterion, 1956). Its first appearance whole in English was in the Carnap jubilee issue of *Synthèse* (Volume 12, 1960), which was subsequently reissued as a book: B. H. Kazemier and D. Vuysje, eds., *Logic and Language* (Dordrecht, Holland: D. Reidel Publishing Co., 1962).

Reprinted with the kind permission of the author, editor, and publisher from *The Philosophy of Rudolf Carnap*, P. A. Schilpp, ed., Library of Living Philosophers, Open Court 1963.

(Bridgman 1934), in much the spirit in which the empiricists of Vienna and elsewhere have inveighed against metaphysics.

What now of the empiricist who would grant certainty to logic, and to the whole of mathematics, and yet would make a clean sweep of other non-empirical theories under the name of metaphysics? The Viennese solution of this nice problem was predicated on language. Metaphysics was meaningless through misuse of language; logic was certain through tautologous use of language.

As an answer to the question "How is logical certainty possible?" this linguistic doctrine of logical truth has its attractions. For there can be no doubt that sheer verbal usage is in general a major determinant of truth. Even so factual a sentence as 'Brutus killed Caesar' owes its truth not only to the killing but equally to our using the component words as we do. Why then should a logically true sentence on the same topic, e.g., 'Brutus killed Caesar or did not kill Caesar', not be said to owe its truth *purely* to the fact that we use our words (in this case 'or' and 'not') as we do? – for it depends not at all for its truth upon the killing.

The suggestion is not, of course, that the logically true sentence is a contingent truth about verbal usage; but rather that it is a sentence which, given the language, automatically becomes true, whereas 'Brutus killed Caesar', given the language, becomes true only contingently on the alleged killing.

Further plausibility accrues to the linguistic doctrine of logical truth when we reflect on the question of alternative logics. Suppose someone puts forward and uses a consistent logic the principles of which are contrary to our own. We are then clearly free to say that he is merely using the familiar particles 'and', 'all', or whatever, in other than the familiar senses, and hence that no real contrariety is present after all. There may of course still be an important failure of intertranslatability, in that the behavior of certain of our logical particles is incapable of being duplicated by paraphrases in his system or vice versa. If the translation in this sense is possible, from his system into ours, then we are pretty sure to protest that he was wantonly using the familiar particles 'and' and 'all' (say) where we might unmisleadingly have used such-and-such other familiar phrasing. This reflection goes to support the view that the truths of logic have no content over and above the meanings they confer on the logical vocabulary.

Much the same point can be brought out by a caricature of a doctrine of Lévy-Bruhl, according to which there are pre-logical peoples who accept certain simple self-contradictions as true. Oversimplifying, no doubt, let us suppose it claimed that these natives accept as true a certain sentence of the form 'p and not p'. Or – not to oversimplify too much –

that they accept as true a certain heathen sentence of the form 'q ka bu q' the English translation of which has the form 'p and not p'. But now just how good a translation is this, and what may the lexicographer's method have been? If any evidence can count against a lexicographer's adoption of 'and' and 'not' as translations of 'ka' and 'bu', certainly the natives' acceptance of 'q ka bu q' as true counts overwhelmingly. We are left with the meaninglessness of the doctrine of there being pre-logical peoples; pre-logicality is a trait injected by bad translators. This is one more illustration of the inseparability of the truths of logic from the meanings of the logical vocabulary.

We thus see that there is something to be said for the naturalness of the linguistic doctrine of logical truth. But before we can get much further we shall have to become more explicit concerning our subject matter.

II

Without thought of any epistemological doctrine, either the linguistic doctrine or another, we may mark out the intended scope of the term 'logical truth', within that of the broader term 'truth', in the following way. First we suppose indicated, by enumeration if not otherwise, what words are to be called logical words; typical ones are 'or', 'not', 'if', 'then', 'and', 'all', 'every', 'only', 'some'. The logical truths, then, are those true sentences which involve only logical words *essentially*. What this means is that any other words, though they may also occur in a logical truth (as witness 'Brutus', 'kill', and 'Caesar' in 'Brutus killed or did not kill Caesar'), can be varied at will without engendering falsity.[1]

Though formulated with reference to language, the above clarification does not of itself hint that logical truths owe their truth to language. What we have thus far is only a delimitation of the class, *per accidens* if you please. Afterward the linguistic doctrine of logical truth, which is an epistemological doctrine, goes on to say that logical truths are true by virtue purely of the intended meanings, or intended usage, of the logical words. Obviously if logical truths *are* true by virtue purely of language, the logical words are the only part of the language that can be concerned in the matter; for these are the only ones that occur essentially.

[1]Substantially this formulation is traced back a century and a quarter, by Bar-Hillel, to Bolzano. But note that the formulation fails of its purpose unless the phrase "can be varied at will," above, is understood to provide for varying the words not only singly but also two or more at a time. E.g., the sentence 'If some men are angels some animals are angels' can be turned into a falsehood by simultaneous substitution for 'men' and 'angels', but not by any substitution for 'angels' alone, nor for 'men', nor for 'animals' (granted the non-existence of angels). For this observation and illustration I am indebted to John R. Myhill, who expresses some indebtedness in turn to Benson Mates. – I added most of this footnote in May 1955, a year after the rest of the essay left my hands.

Elementary logic, as commonly systematized nowadays, comprises truth-function theory, quantification theory, and identity theory. The logical vocabulary for this part, as commonly rendered for technical purposes, consists of truth-function signs (corresponding to 'or', 'and', 'not', etc.), quantifiers and their variables, and '='.

The further part of logic is set theory, which requires there to be classes among the values of its variables of quantification. The one sign needed in set theory, beyond those appropriate to elementary logic, is the connective '∈' of membership. Additional signs, though commonly used for convenience, can be eliminated in well-known ways.

In this dichotomy I leave metatheory, or logical syntax, out of account. For, either it treats of special objects of an extralogical kind, viz., notational expressions, or else, if these are made to give way to numbers by arithmetization, it is reducible via number theory to set theory.

I will not here review the important contrasts between elementary logic and set theory, except for the following one. Every truth of elementary logic is obvious (whatever this really means), or can be made so by some series of individually obvious steps. Set theory, in its present state anyway, is otherwise. I am not alluding here to Gödel's incompleteness principle, but to something right on the surface. Set theory was straining at the leash of intuition ever since Cantor discovered the higher infinites; and with the added impetus of the paradoxes of set theory the leash was snapped. Comparative set theory has now long been the trend; for, so far as is known, no consistent set theory is both adequate to the purposes envisaged for set theory and capable of substantiation by steps of obvious reasoning from obviously true principles. What we do is develop one or another set theory by obvious reasoning, or elementary logic, from unobvious first principles which are set down, whether for good or for the time being, by something very like convention.

Altogether, the contrasts between elementary logic and set theory are so fundamental that one might well limit the word 'logic' to the former (though I shall not), and speak of set theory as mathematics in a sense exclusive of logic. To adopt this course is merely to deprive '∈' of the status of a logical word. Frege's derivation of arithmetic would then cease to count as a derivation from logic; for he used set theory. At any rate we should be prepared to find that the linguistic doctrine of logical truth holds for elementary logic and fails for set theory, or vice versa. Kant's readiness to see logic as analytic and arithmetic as synthetic, in particular, is not superseded by Frege's work (as Frege supposed; see Frege 1950: sections 87f., 109) if 'logic' be taken as elementary logic. And for Kant logic certainly did not include set theory.

III

Where someone disagrees with us as to the truth of a sentence, it often happens that we can convince him by getting the sentence from other sentences, which he does accept, by a series of steps each of which he accepts. Disagreement which cannot be thus resolved I shall call *deductively irresoluble.* Now if we try to warp the linguistic doctrine of logical truth around into something like an experimental thesis, perhaps a first approximation will run thus: *Deductively irresoluble disagreement as to a logical truth is evidence of deviation in usage* (*or meanings*) *of words.* This is not yet experimentally phrased, since one term of the affirmed relationship, viz., 'usage' (or 'meanings'), is in dire need of an independent criterion. However, the formulation would seem to be fair enough within its limits; so let us go ahead with it, not seeking more subtlety until need arises.

Already the obviousness or potential obviousness of elementary logic can be seen to present an insuperable obstacle to our assigning any experimental meaning to the linguistic doctrine of elementary logical truth. Deductively irresoluble dissent from an elementary logical truth *would* count as evidence of deviation over meanings if anything can, but simply because dissent from a logical truism is as extreme as dissent can get.

The philosopher, like the beginner in algebra, works in danger of finding that his solution-in-progress reduces to '0=0'. Such is the threat to the linguistic theory of elementary logical truth. For, that theory now seems to imply nothing that is not already implied by the fact that elementary logic is obvious or can be resolved into obvious steps.

The considerations which were adduced in §I, to show the naturalness of the linguistic doctrine, are likewise seen to be empty when scrutinized in the present spirit. One was the circumstance that alternative logics are inseparable practically from mere change in usage of logical words. Another was that illogical cultures are indistinguishable from ill-translated ones. But both of these circumstances are adequately accounted for by mere obviousness of logical principles, without help of a linguistic doctrine of logical truth. For, there can be no stronger evidence of a change in usage than the repudiation of what had been obvious, and no stronger evidence of bad translation than that it translates earnest affirmations into obvious falsehoods.

Another point in §I was that true sentences generally depend for their truth on the traits of their language in addition to the traits of their subject matter; and that logical truths then fit neatly in as the limiting case where the dependence on traits of the subject matter is nil. Consider, however, the logical truth 'Everything is self-identical', or '$(x)(x=x)$'. We *can* say that it depends for its truth on traits of the language (specifically on the usage of '='), and not on traits of its subject matter; but we can also say, alternatively, that it depends on an obvious trait, viz., self-identity, of its subject matter, viz., everything. The tendency of our present reflections is that there is no difference.

I have been using the vaguely psychological word "obvious" non-technically, assigning it no explanatory value. My suggestion is merely that the linguistic doctrine of elementary logical truth likewise leaves explanation unbegun. I do not suggest that the linguistic doctrine is false and some doctrine of ultimate and inexplicable insight into the obvious traits of reality is true, but only that there is no real difference between these two pseudodoctrines.

Turning away now from elementary logic, let us see how the linguistic doctrine of logical truth fares in application to set theory. As noted in §II, we may think of '∈' as the one sign for set theory in addition to those of elementary logic. Accordingly the version of the linguistic doctrine which was italicized at the beginning of the present section becomes, in application to set theory, this: Among persons who are already in agreement on elementary logic, any deductively irresoluble disagreement as to a truth of set theory is evidence of deviation in usage (or meaning) of '∈'.

This thesis is not trivial in quite the way in which the parallel thesis for elementary logic was seen to be. It is not indeed experimentally significant as it stands, simply because of the lack, noted earlier, of a separate criterion for usage or meaning. But it does seem reasonable, by the following reasoning.

Any acceptable evidence of usage or meaning of words must reside surely either in the observable circumstances under which the words are uttered (in the case of concrete terms referring to observable individuals) or in the affirmation and denial of sentences in which the words occur. Only the second alternative is relevant to '∈'. Therefore any evidence of deviation in usage or meaning of '∈' must reside in disagreement on sentences containing '∈'. This is not, of course, to say of *every* sentence containing '∈' that disagreement over it establishes deviation in usage or meaning of '∈'. We have to assume in the first place that the speaker under investigation agrees with us on the meanings of words other than '∈' in the sentences in question. And it might well be that, even from among the sentences containing only '∈' and words on whose meanings he agrees with us, there is only a select species S which is so fundamental that he cannot dissent from them without betraying deviation in his

usage or meaning of '∈'. But S may be expected surely to include some (if not all) of the sentences which contain *nothing* but '∈' and the elementary logical particles; for it is these sentences, insofar as true, that constitute (pure, or unapplied) set theory. But it is difficult to conceive of how to be other than democratic toward the truths of set theory. In exposition we may select some of these truths as so-called postulates and deduce others from them, but this is subjective discrimination, variable at will, expository and not set-theoretic. We do not change our meaning of '∈' between the page where we show that one particular truth is deducible by elementary logic from another and the page where we show the converse. Given this democratic outlook, finally, the law of sufficient reason leads us to look upon S as including *all* the sentences which contain only '∈' and the elementary logical particles. It then follows that anyone in agreement on elementary logic and in irresoluble disagreement on set theory is in deviation with respect to the usage or meaning of '∈'; and this was the thesis.

The effect of our effort to inject content into the linguistic doctrine of logical truth has been, up to now, to suggest that the doctrine says nothing worth saying about elementary logical truth, but that when applied to set-theoretic truth it makes for a reasonable partial condensation of the otherwise vaporous notion of meaning as applied to '∈'.

IV

The linguistic doctrine of logical truth is sometimes expressed by saying that such truths are true by linguistic convention. Now if this be so, certainly the conventions are not in general explicit. Relatively few persons, before the time of Carnap, had ever seen any convention that engendered truths of elementary logic. Nor can this circumstance be ascribed merely to the slipshod ways of our predecessors. For it is impossible in principle, even in an ideal state, to get even the most elementary part of logic exclusively by the explicit application of conventions stated in advance. The difficulty is the vicious regress, familiar from Lewis Carroll, which I have elaborated elsewhere (1936) [reprinted in this volume]. Briefly the point is that the logical truths, being infinite in number, must be given by general conventions rather than singly; and logic is needed then to begin with, in the metatheory, in order to apply the general conventions to individual cases.

"In dropping the attributes of deliberateness and explicitness from the notion of linguistic convention," I went on to complain in the aforementioned paper, "we risk depriving the latter of any explanatory force and reducing it to an idle label." It would seem that to call elementary logic true by convention is to add nothing but a metaphor to the linguistic doctrine of logical truth which, as applied to elementary logic, has itself come to seem rather an empty figure (cf. §III).

The case of set theory, however, is different on both counts. For set theory the linguistic doctrine has seemed less empty (cf. §III); in set theory, moreover, convention in quite the ordinary sense seems to be pretty much what goes on (cf. §II). Conventionalism has a serious claim to attention in the philosophy of mathematics, if only because of set theory. Historically, though, conventionalism was encouraged in the philosophy of mathematics rather by the non-Euclidean geometries and abstract algebras, with little good reason. We can contribute to subsequent purposes by surveying this situation. Further talk of set theory is deferred to §V.

In the beginning there was Euclidean geometry, a compendium of truths about form and void; and its truths were not based on convention (except as a conventionalist might, begging the present question, apply this tag to everything mathematical). Its truths were in practice presented by deduction from so-called postulates (including axioms; I shall not distinguish); and the selection of truths for this role of postulate, out of the totality of truths of Euclidean geometry, was indeed a matter of convention. But this is not *truth* by convention. The truths were there, and what was conventional was merely the separation of them into those to be taken as starting point (for purposes of the exposition at hand) and those to be deduced from them.

The non-Euclidean geometries came of artificial deviations from Euclid's postulates, without thought (to begin with) of true interpretation. These departures were doubly conventional; for Euclid's postulates were a conventional selection from among the truths of geometry, and then the departures were arbitrarily or conventionally devised in turn. But still there was no truth by convention, because there was no truth.

Playing within a non-Euclidean geometry, one might conveniently make believe that his theorems were interpreted and true; but even such conventional make-believe is not truth by convention. For it is not really truth at all; and what is conventionally pretended is that the theorems are true by non-convention.

Non-Euclidean geometries have, in the fullness of time, received serious interpretations. This means that ways have been found of so construing the hitherto unconstrued terms as to identify the at first conventionally chosen set of non-sentences with some genuine truths, and truths presumably not by convention. The status of an interpreted non-Euclidean geometry differs in no basic way from the original status of Euclidean geometry, noted above.

Uninterpreted systems became quite the fashion after the advent of non-Euclidean geometries. This fashion helped to cause, and was in turn encouraged by, an increasingly formal approach to mathematics. Methods had to become more formal to make up for the unavailability, in uninterpreted systems, of intuition. Conversely, disinterpretation served as a crude but useful device (until Frege's syntactical approach came to be appreciated) for achieving formal rigor uncorrupted by intuition.

The tendency to look upon non-Euclidean geometries as true by convention applied to uninterpreted systems generally, and then carried over from these to mathematical systems generally. A tendency indeed developed to look upon all mathematical systems as, qua mathematical, uninterpreted. This tendency can be accounted for by the increase of formality, together with the use of disinterpretation as a heuristic aid to formalization. Finally, in an effort to make some sense of mathematics thus drained of all interpretation, recourse was had to the shocking quibble of identifying mathematics merely with the elementary logic which leads from uninterpreted postulates to uninterpreted theorems (see my 1936: section I) [pp. 329–38 in this volume]. What is shocking about this is that it puts arithmetic qua interpreted theory of number, and analysis qua interpreted theory of functions, and geometry qua interpreted theory of space, outside mathematics altogether.

The substantive reduction of mathematics to logic by Frege, Whitehead, and Russell is of course quite another thing. It is a reduction not to elementary logic but to set theory; and it is a reduction of genuine interpreted mathematics, from arithmetic onward.

V

Let us then put aside these confusions and get back to set theory. Set theory is pursued as interpreted mathematics, like arithmetic and analysis; indeed, it is to set theory that those further branches are reducible. In set theory we discourse about certain immaterial entities, real or erroneously alleged, viz., sets, or classes. And it is in the effort to make up our minds about genuine truth and falsity of sentences about these objects that we find ourselves engaged in something very like convention in an ordinary non-metaphorical sense of the word. We find ourselves making deliberate choices and setting them forth unaccompanied by any attempt at justification other than in terms of elegance and convenience. These adoptions, called postulates, and their logical consequences (via elementary logic), are true until further notice.

So here is a case where postulation can plausibly be looked on as constituting truth by convention. But in §IV we have seen how the philosophy of mathematics can be corrupted by supposing that postulates always play that role. Insofar as we would epistemologize and not just mathematize, we might divide postulation as follows. Uninterpreted postulates may be put aside, as no longer concerning us; and on the interpreted side we may distinguish between *legislative* and *discursive* postulation. Legislative postulation institutes truth by convention, and seems plausibly illustrated in contemporary set theory. On the other hand discursive postulation is mere selection, from a pre-existing body of truths, of certain ones for use as a basis from which to derive others, initially known or unknown. What discursive postulation fixes is not truth, but only some particular ordering of the truths, for purposes perhaps of pedagogy or perhaps of inquiry into logical relationships (logical in the sense of elementary logic). All postulation is of course conventional, but only legislative postulation properly hints of *truth* by convention.

It is well to recognize, if only for its distinctness, yet a further way in which convention can enter; viz., in the adoption of new notations for old ones, without, as one tends to say, change of theory. Truths containing the new notation are conventional transcriptions of sentences true apart from the convention in question. They depend for their truth partly on language, but then so did 'Brutus killed Caesar' (cf. §I). They come into being through a conventional adoption of a new sign, and they become true through conventional definition of that sign *together with* whatever made the corresponding sentences in the old notation true.

Definition, in a properly narrow sense of the word, is convention in a properly narrow sense of the word. But the phrase 'true by definition' must be taken cautiously; in its strictest usage it refers to a transcription, by the definition, of a truth of elementary logic. Whether such a sentence is true by convention depends on whether the logical truths themselves be reckoned as true by convention. Even an outright equation or biconditional connection of the definiens and the definiendum is a definitional transcription of a prior logical truth of the form '$x=x$' or '$p \equiv p$'.

Definition commonly so-called is not thus narrowly conceived, and must for present purposes be divided, as postulation was divided, into legislative and discursive. Legislative definition introduces a notation hitherto unused, or used only at variance with the practice proposed, or used also at variance, so that a convention is wanted to settle the ambiguity. Discursive definition, on the other hand, sets forth a pre-existing relation of interchangeability or coextensiveness between notations in already familiar usage. A frequent purpose of this activity is to show how

some chosen part of language can be made to serve the purposes of a wider part. Another frequent purpose is language instruction.

It is only legislative definition, and not discursive definition or discursive postulation, that makes a conventional contribution to the truth of sentences. Legislative postulation, finally, affords truth by convention unalloyed.

Increasingly the word 'definition' connotes the formulas of definition which appear in connection with formal systems, signaled by some extra-systematic sign such as '$=_{df}$'. Such definitions are best looked upon as correlating two systems, two notations, one of which is prized for its economical lexicon and the other for its brevity or familiarity of expression (see my 1953a: 26f.). Definitions so used can be either legislative or discursive in their inception. But this distinction is in practice left unindicated, and wisely; for it is a distinction only between particular acts of definition, and not germane to the definition as an enduring channel of intertranslation.

The distinction between the legislative and the discursive refers thus to the act, and not to its enduring consequence, in the case of postulation as in the case of definition. This is because we are taking the notion of truth by convention fairly literally and simple-mindedly, for lack of an intelligible alternative. So conceived, conventionality is a passing trait, significant at the moving front of science but useless in classifying the sentences behind the lines. It is a trait of events and not of sentences.

Might we not still project a derivative trait upon the sentences themselves, thus speaking of a sentence as forever true by convention if its first adoption as true was a convention? No; this, if done seriously, involves us in the most unrewarding historical conjecture. Legislative postulation contributes truths which become integral to the corpus of truths; the artificiality of their origin does not linger as a localized quality, but suffuses the corpus. If a subsequent expositor singles out those once legislatively postulated truths again as postulates, this signifies nothing; he is engaged only in discursive postulation. He could as well choose his postulates from elsewhere in the corpus, and will if he thinks this serves his expository ends.

VI

Set theory, currently so caught up in legislative postulation, may some day gain a norm – even a strain of obviousness, perhaps – and lose all trace of the conventions in its history. A day could likewise have been when our elementary logic was itself instituted as a deliberately conventional deviation from something earlier, instead of evolving, as it did, mainly by unplanned shifts of form and emphasis coupled with casual novelties of notation.

Today indeed there are dissident logicians even at the elementary level, propounding deviations from the law of the excluded middle. These deviations, insofar as meant for serious use and not just as uninterpreted systems, are as clear cases of legislative postulation as the ones in set theory. For here we have again, quite as in set theory, the propounding of a deliberate choice unaccompanied (conceivably) by any attempt at justification other than in terms of convenience.

This example from elementary logic controverts no conclusion we have reached. According to §§I and III, the departure from the law of the excluded middle would count as evidence of revised usage of 'or' and 'not'. (This judgment was upheld in §III, though disqualified as evidence for the linguistic doctrine of logical truth.) For the deviating logician the words 'or' and 'not' are unfamiliar, or defamiliarized; and his decisions regarding truth values for their proposed contexts can then be just as genuinely a matter of deliberate convention as the decisions of the creative set theorist regarding contexts of '$\in$'.

The two cases are indeed much alike. Not only is departure from the classical logic of 'or' and 'not' evidence of revised usage of 'or' and 'not'; likewise, as argued at length in §III, divergences between set theorists may reasonably be reckoned to revised usage of '$\in$'. Any such revised usage is conspicuously a matter of convention, and can be declared by legislative postulation.

We have been at a loss to give substance to the linguistic doctrine, particularly of elementary logical truth, or to the doctrine that the familiar truths of logic are true by convention. We have found some sense in the notion of truth by convention, but only as attaching to a process of adoption, viz., legislative postulation, and not as a significant lingering trait of the legislatively postulated sentence. Surveying current events, we note legislative postulation in set theory and, at a more elementary level, in connection with the law of the excluded middle.

And do we not find the same continually in the theoretical hypotheses of natural science itself? What seemed to smack of convention in set theory (§V), at any rate, was "deliberate choice, set forth unaccompanied by any attempt at justification other than in terms of elegance and convenience"; and to what theoretical hypothesis of natural science might not this same character be attributed? For surely the justification of any theoretical hypothesis can, at the time of hypothesis, consist in no more than the elegance or convenience which the hypothesis brings to the containing body of laws and data. How then are we to delimit the category of legislative postulation, short of including under it every new act of scientific hypothesis?

The situation may seem to be saved, for ordinary hypotheses in natural science, by there being some indirect but eventual confrontation with empirical data. However, this confrontation can be remote; and, conversely, some such remote confrontation with experience may be claimed even for pure mathematics and elementary logic. The semblance of a difference in this respect is largely due to overemphasis of departmental boundaries. For a self-contained theory which we can check with experience includes, in point of fact, not only its various theoretical hypotheses of so-called natural science but also such portions of logic and mathematics as it makes use of. Hence I do not see how a line is to be drawn between hypotheses which confer truth by convention and hypotheses which do not, short of reckoning *all* hypotheses to the former category save perhaps those actually derivable or refutable by elementary logic from what Carnap used to call protocol sentences. But this version, besides depending to an unwelcome degree on the debatable notion of protocol sentences, is far too inclusive to suit anyone.

Evidently our troubles are waxing. We had been trying to make sense of the role of convention in a priori knowledge. Now the very distinction between a priori and empirical begins to waver and dissolve, at least as a distinction between sentences. (It could of course still hold as a distinction between factors in one's adoption of a sentence, but both factors might be operative everywhere.)

VII

Whatever our difficulties over the relevant distinctions, it must be conceded that logic and mathematics do seem qualitatively different from the rest of science. Logic and mathematics hold conspicuously aloof from any express appeal, certainly, to observation and experiment. Having thus nothing external to look to, logicians and mathematicians look closely to notation and explicit notational operations: to expressions, terms, substitution, transposition, cancellation, clearing of fractions, and the like. This concern of logicians and mathematicians with syntax (as Carnap calls it) is perennial, but in modern times it has become increasingly searching and explicit, and has even prompted, as we see, a linguistic philosophy of logical and mathematical truth.

On the other hand an effect of these same formal developments in modern logic, curiously, has been to show how to divorce mathematics (other than elementary logic) from any peculiarly notational considerations not equally relevant to natural science. By this I mean that mathematics can be handled (insofar as it can be handled at all) by axiomatization, outwardly quite like any system of hypotheses elsewhere in science; and elementary logic can then be left to extract the theorems.

The consequent affinity between mathematics and systematized natural science was recognized by Carnap when he propounded his P-rules alongside his L-rules or meaning postulates. Yet he did not look upon the P-rules as engendering analytic sentences, sentences true purely by language. How to sustain this distinction has been very much our problem in these pages, and one on which we have found little encouragement.

Carnap appreciated this problem, in *Logical Syntax*, as a problem of finding a difference in kind between the P-rules (or the truths thereby specified) and the L-rules (or the L-truths, analytic sentences, thereby specified). Moreover he proposed an ingenious solution (1937; section 50). In effect he characterized the logical (including mathematical) vocabulary as the largest vocabulary such that (1) there are sentences which contain only that vocabulary and (2) all such sentences are determinable as true or false by a purely syntactical condition – i.e., by a condition which speaks only of concatenation of marks. Then he limited the L-truths in effect to those involving just the logical vocabulary essentially.[2]

Truths given by P-rules were supposedly excluded from the category of logical truth under this criterion, because, though the rules specifying them are formally stated, the vocabulary involved can also be recombined to give sentences whose truth values are not determinate under any set of rules formally formulable in advance.

At this point one can object (pending a further expedient of Carnap's, which I shall next explain) that the criterion based on (1) and (2) fails of its purpose. For, consider to begin with the totality of those sentences which are expressed purely within what Carnap (or anyone) would want to count as logical (and mathematical) vocabulary. Suppose, in conformity with (2), that the division of this totality into the true and the false is reproducible in purely syntactical terms. Now surely the adding of one general term of an extra-logical kind, say 'heavier than', is not going to alter the situation. The truths which are expressible in terms of just 'heavier than', together with the logical vocabulary, will be truths of only the most general kind, such as '$(\exists x)(\exists y)(x$ is heavier than $y)$', '$(x) \sim (x$ is heavier than $x)$', and '$(x)(y)(z)(x$ is heavier than $y . y$ is heavier than $z . \supset . x$ is heavier than $z)$'. The division of the truths from

[2]Cf. §I above. Also, for certain reservations conveniently postponed at the moment, see §IX on "essential predication."

the falsehoods in this supplementary domain can probably be reproduced in syntactical terms if the division of the original totality could. But then, under the criterion based on (1) and (2), 'heavier than' qualifies for the logical vocabulary. And it is hard to see what whole collection of general terms of natural science might not qualify likewise.

The further expedient, by which Carnap met this difficulty, was his use of Cartesian co-ordinates (1937: sections 3, 15). Under this procedure, each spatio-temporal particular c becomes associated with a class K of quadruples of real numbers, viz., the class of those quadruples which are the co-ordinates of component point events of c. Further let us write $K[t]$ for the class of triples which with t appended belong to K; thus $K[t]$ is that class of triples of real numbers which is associated with the momentary state of object c at time t. Then, in order to say, e.g., that c_1 is heavier than c_2 at time t, we say '$H(K_1[t], K_2[t])$', which might be translated as 'The momentary object associated with $K_1[t]$ is heavier than that associated with $K_2[t]$'. Now $K_1[t]$ and $K_2[t]$ are, in every particular case, purely mathematical objects; viz., classes of triples of real numbers. So let us consider all the true and false sentences of the form '$H(K_1[t], K_2[t])$' where, in place of '$K_1[t]$' and '$K_2[t]$', we have purely logico-mathematical designations of particular classes of triples of real numbers. There is no reason to suppose that all the truths of *this* domain can be exactly segregated in purely syntactical terms. Thus inclusion of 'H' does violate (2), and therefore 'H' fails to qualify as logical vocabulary. By adhering to the method of co-ordinates and thus reconstruing all predicates of natural science in the manner here illustrated by 'H', Carnap overcomes the objection noted in the preceding paragraph.

To sum up very roughly, this theory characterizes logic (and mathematics) as the largest part of science within which the true–false dichotomy *can* be reproduced in syntactical terms. This version may seem rather thinner than the claim that logic and mathematics are somehow true by linguistic convention, but at any rate it is more intelligible, and, if true, perhaps interesting and important. To become sure of its truth, interest, and importance, however, we must look more closely at this term 'syntax'.

As used in the passage: "The terms 'sentence' and 'direct consequence' are the two primitive terms of logical syntax" (Carnap 1935c: 47), the term 'syntax' is of course irrelevant to a thesis. The relevant sense is that rather in which it connotes discourse about marks and their succession. But here still we must distinguish degrees of inclusiveness; two different degrees are exemplified in *Logical Syntax*, according as the object language is Carnap's highly restricted Language I or his more powerful Language II. For the former, Carnap's formulation of logical truth is narrowly syntactical in the manner of familiar formalizations of logical systems by axioms and rules of inference. But Gödel's proof of the incompletability of elementary number theory shows that no such approach can be adequate to mathematics in general, nor in particular to set theory, nor to Language II. For Language II, in consequence, Carnap's formulation of logical truth proceeded along the lines rather of Tarski's technique of truth definition.[3] The result was still a purely syntactical specification of the logical truths, but only in this more liberal sense of 'syntactical': it was couched in a vocabulary consisting (in effect) of (a) names of signs, (b) an operator expressing concatenation of expressions, and (c), by way of auxiliary machinery, the whole logical (and mathematical) vocabulary itself.

So construed, however, the thesis that logico-mathematical truth is syntactically specifiable becomes uninteresting. For, what it says is that logico-mathematical truth is specifiable in a notation consisting solely of (a), (b), *and* the whole logico-mathematical vocabulary itself. But *this* thesis would hold equally if 'logico-mathematical' were broadened (at *both* places in the thesis) to include physics, economics, and anything else under the sun; Tarski's routine of truth definition would still carry through just as well. No special trait of logic and mathematics has been singled out after all.

Strictly speaking, the position is weaker still. The mathematics appealed to in (c) must, as Tarski shows, be a yet more inclusive mathematical theory in certain respects than that for which truth is being defined. It was largely because of his increasing concern over this self-stultifying situation that Carnap relaxed his stress on syntax, in the years following *Logical Syntax*, in favor of semantics.

VIII

Even if logical truth were specifiable in syntactical terms, this would not show that it was grounded in language. Any *finite* class of truths (to take an extreme example) is clearly reproducible by a membership condition couched in as narrowly syntactical terms as you please; yet we certainly cannot say of every finite class of truths that its members are true purely by language. Thus the ill-starred doctrine of syntactical specifiability of logical truth was always something other than the linguistic doctrine of

[3]*Logical Syntax*, especially §§34a–i, 60a–d, 71a–d. These sections had been omitted from the German edition, but only for lack of space; cf. p. xi of the English edition. Meanwhile they had appeared as articles: "die Antinomien ..." and "ein Gültigkeitskriterium ..." At that time Carnap had had only partial access to Tarski's ideas (cf. "Gültigkeitskriterium," footnote 3), the full details of which reached the non-Slavic world in 1935 in Tarski's *Wahrheitsbegriff*."

logical truth, if this be conceived as the doctrine that logical truth is grounded in language. In any event the doctrine of syntactical specifiability, which we found pleasure in being able to make comparatively clear sense of, has unhappily had to go by the board. The linguistic doctrine of logical truth, on the other hand, goes sturdily on.

The notion of logical truth is now counted by Carnap as semantical. This of course does not of itself mean that logical truth is gounded in language; for note that the general notion of truth is also semantical, though truth in general is not grounded purely in language. But the semantical attribute of logical truth, in particular, *is* one which, according to Carnap, is grounded in language: in convention, fiat, meaning. Such support as he hints for this doctrine, aside from ground covered in §§I–VI, seems to depend on an analogy with what goes on in the propounding of artificial languages; and I shall now try to show why I think the analogy mistaken.

I may best schematize the point by considering a case, not directly concerned with logical truth, where one might typically produce an artificial language as a step in an argument. This is the imaginary case of a logical positivist, say Ixmann, who is out to defend scientists against the demands of a metaphysician. The metaphysician argues that science presupposes metaphysical principles, or raises metaphysical problems, and that the scientists should therefore show due concern. Ixmann's answer consists in showing in detail how people (on Mars, say) might speak a language quite adequate to all of our science but, unlike our language, incapable of expressing the alleged metaphysical issues. (I applaud this answer, and think it embodies the most telling component of Carnap's own anti-metaphysical representations; but here I digress.) Now how does our hypothetical Ixmann specify that doubly hypothetical language? By telling us, at least to the extent needed for his argument, what these Martians are to be imagined as uttering and what they are thereby to be understood to mean. Here is Carnap's familiar duality of formation rules and transformation rules (or meaning postulates), as rules of language. But these rules are part only of Ixmann's narrative machinery, not part of what he is portraying. He is not representing his hypothetical Martians themselves as somehow explicit on formation and transformation rules. Nor is he representing there to be any intrinsic difference between those truths which happen to be disclosed to us by his partial specifications (his transformation rules) and those further truths, hypothetically likewise known to the Martians of his parable, which he did not trouble to sketch in.

The threat of fallacy lurks in the fact that Ixmann's rules are indeed arbitrary fiats, as is his whole Martian parable. The fallacy consists in confusing levels, projecting the conventional character of the rules into the story, and so misconstruing Ixmann's parable as attributing truth legislation to his hypothetical Martians.

The case of a non-hypothetical artificial language is in principle the same. Being a new invention, the language has to be explained; and the explanation will proceed by what may certainly be called formation and transformation rules. These rules will hold by arbitrary fiat, the artifex being boss. But all we can reasonably ask of these rules is that they enable us to find corresponding to each of his sentences a sentence of like truth value in familiar ordinary language. There is no (to me) intelligible additional decree that we can demand of him as to the boundary between analytic and synthetic, logic and fact, among his truths. We may well decide to extend our word 'analytic' or 'logically true' to sentences of his language which he in his explanations has paired off fairly directly with English sentences so classified by us; but this is our decree, regarding our word 'analytic' or 'logically true'.

IX

We had in §II to form some rough idea of what logical truth was supposed to take in, before we could get on with the linguistic doctrine of logical truth. This we did, with help of the general notion of truth[4] together with a partial enumeration of the logical vocabulary of a particular language. In §VII we found hope of a less provincial and accidental characterization of logical vocabulary; but it failed. Still, the position is not intolerable. We well know from modern logic how to devise a technical notation which is admirably suited to the business of 'or', 'not', 'and', 'all', 'only', and such other particles as we would care to count as logical; and to enumerate the signs and constructions of that technical notation, or a theoretically adequate subset of them, is the work of a moment (cf. §II). Insofar as we are content to think of all science as fitted within that stereotyped logical framework – and there is no hardship in so doing – our notion of logical vocabulary is precise. And so, derivatively, is our notion of logical truth. But only in point of extent. There is no epistemological corollary as to the *ground* of logical truth (cf. §II).

Even this halfway tolerable situation obtains only for logical truth in a relatively narrow sense, omitting truths by "essential predication" (Aristotle) such as 'No bachelor is married'. I tend to reserve the term 'logically true' for the narrower domain, and to use the term 'analytic' for the more inclusive domain which includes truths by essential predication. Carnap on the contrary has used both terms in the broader sense. But the problems of the two subdivisions of the analytic class differ in such a way that it has been convenient up to now in this essay to treat mainly of logical truth in the narrower sense.

[4]In defense of this general notion, in invidious contrast to that of analyticity, see my *From a Logical Point of View*, pp. 137f.

The truths by essential predication are sentences which can be turned into logical truths by supplanting certain simple predicates (e.g., 'bachelor') by complex synonyms (e.g., 'man not married'). This formulation is not inadequate to such further examples as 'if *A* is part of *B* and *B* is part of *C* then *A* is part of *C*'; this case can be managed by using for 'is part of' the synonym 'overlaps nothing save what overlaps' (after Goodman 1951). The relevant notion of synonymy is simply *analytic* coextensiveness (however circular this might be as a definition).

To count analyticity a genus of logical truth is to grant, it may seem, the linguistic doctrine of logical truth; for the term 'analytic' directly suggests truth by language. But this suggestion can be adjusted, in parallel to what was said of 'true by definition' in §V. 'Analytic' means true by synonymy and logic, hence no doubt true by language and logic, and simply true by language *if* the linguistic doctrine of logical truth is right. Logic itself, throughout these remarks, may be taken as including or excluding set theory (and hence mathematics), depending on further details of one's position.

What has made it so difficult for us to make satisfactory sense of the linguistic doctrine is the obscurity of 'true by language'. Now 'synonymous' lies within that same central obscurity; for, about the best we can say of synonymous predicates is that they are somehow "co-extensive by language." The obscurity extends, of course, to 'analytic'.

One quickly identifies certain seemingly transparent cases of synonymy, such as 'bachelor' and 'man not married', and senses the triviality of associated sentences such as 'No bachelor is married'. Conceivably the mechanism of such recognition, when better understood, might be made the basis of a definition of synonymy and analyticity in terms of linguistic behavior. On the other hand such an approach might make sense only of something like degrees of synonymy and analyticity. I see no reason to expect that the full-width analyticity which Carnap and others make such heavy demands upon can be fitted to such a foundation in even an approximate way. In any event, we at present lack any tenable general suggestion, either rough and practical or remotely theoretical, as to what it is to be an analytic sentence. All we have are purported illustrations, and claims that the truths of elementary logic, with or without the rest of mathematics, should be counted in. Wherever there has been a semblance of a general criterion, to my knowledge, there has been either some drastic failure such as tended to admit all or no sentences as analytic, or there has been a circularity of the kind noted three paragraphs back, or there has been a dependence on terms like 'meaning', 'possible', 'conceivable', and the like, which are at least as mysterious (and in the same way) as what we want to define. I have expatiated on these troubles elsewhere (1951d), as has White.

Logical truth (in my sense, excluding the additional category of essential predication) is, we saw, well enough definable (relatively to a fixed logical notation). *Elementary* logical truth can even be given a narrowly syntactical formulation, such as Carnap once envisaged for logic and mathematics as a whole (cf. §VII); for the deductive system of elementary logic is known to be complete. But when we would supplement the logical truths by the rest of the so-called analytic truths, true by essential predication, then we are no longer able even to say what we are talking about. The distinction itself, and not merely an epistemological question concerning it, is what is then in question.

What of settling the limits of the broad class of analytic truths by fixing on a standard language as we did for logical truth? No, the matter is very different. Once given the logical vocabulary, we have a means of clearly marking off the species logical truth within the genus truth. But the intermediate genus analyticity is not parallel, for it does not consist of the truths which contain just a certain vocabulary essentially (in the sense of §II). To segregate analyticity we should need rather some sort of accounting of synonymies throughout a universal or all-purpose language. No regimented universal language is at hand, however, for consideration; what Carnap has propounded in this direction have of course been only illustrative samples, fragmentary in scope. And even if there were one, it is not clear by what standards we would care to settle questions of synonymy and analyticity within it.

X

Carnap's present position (particularly 1952) is that one has specified a language quite rigorously only when he has fixed, by dint of so-called meaning postulates, what sentences are to count as analytic. The proponent is supposed to distinguish between those of his declarations which count as meaning postulates, and thus engender analyticity, and those which do not. This he does, presumably, by attaching the label 'meaning postulate'.

But the sense of this label is far less clear to me than four causes of its seeming to be clear. Which of these causes has worked on Carnap, if any,

I cannot say; but I have no doubt that all four have worked on his readers. One of these causes is misevaluation of the role of convention in connection with artificial language; thus note the unattributed fallacy described in §VIII. Another is misevaluation of the conventionality of postulates: failure to appreciate that postulates, though they are postulates always by fiat, are not *therefore* true by fiat (cf. §§IV–V). A third is over-estimation of the distinctive nature of postulates, and of definitions, because of conspicuous and peculiar roles which postulates and definitions have played in situations not really relevant to present concerns: postulates in uninterpreted systems (cf. §IV), and definitions in double systems of notation (cf. §V). A fourth is misevaluation of legislative postulation and legislative definition themselves, in two respects: failure to appreciate that this legislative trait is a trait of scientific hypotheses very generally (cf. §VI), and failure to appreciate that it is a trait of the passing event rather than of the truth which is thereby instituted (cf. end of §V).

Suppose a scientist introduces a new term, for a certain substance or force. He introduces it by an act either of legislative definition or of legislative postulation. Progressing, he evolves hypotheses regarding further traits of the named substance or force. Suppose now that some such eventual hypothesis, well attested, identifies this substance or force with one named by a complex term built up of other portions of his scientific vocabulary. We all know that this new identity will figure in the ensuing developments quite on a par with the identity which first came of the act of legislative definition, if any, or on a par with the law which first came of the act of legislative postulation. Revisions, in the course of further progress, can touch any of these affirmations equally. Now I urge that scientists, proceeding thus, are not thereby slurring over any meaningful distinction. Legislative acts occur again and again; on the other hand a dichotomy of the resulting truths themselves into analytic and synthetic, truths by meaning postulate and truths by force of nature, has been given no tolerably clear meaning even as a methodological ideal.

One conspicuous consequence of Carnap's belief in this dichotomy may be seen in his attitude toward philosophical issues as to what there is (Quine 1951a). It is only by assuming the cleavage between analytic and synthetic truths that he is able to declare the problem of universals to be a matter not of theory but of linguistic decision. Now I am as impressed as anyone with the vastness of what language contributes to science and to one's whole view of the world; and in particular I grant that one's hypothesis as to what there is, e.g., as to there being universals, is at bottom just as arbitrary or pragmatic a matter as one's adoption of a new brand of set theory or even a new system of bookkeeping. Carnap in turn recognizes that such decisions, however conventional, "will nevertheless usually be influenced by theoretical knowledge" (1950: §2). But what impresses me more than it does Carnap is how well this whole attitude is suited also to the theoretical hypotheses of natural science itself, and how little basis there is for a distinction.

The lore of our fathers is a fabric of sentences. In our hands it develops and changes, through more or less arbitrary and deliberate revisions and additions of our own, more or less directly occasioned by the continuing stimulation of our sense organs. It is a pale gray lore, black with fact and white with convention. But I have found no substantial reasons for concluding that there are any quite black threads in it, or any white ones.

On the nature of mathematical truth

CARL G. HEMPEL

1. The problem

It is a basic principle of scientific inquiry that no proposition and no theory is to be accepted without adequate grounds. In empirical science, which includes both the natural and the social sciences, the grounds for the acceptance of a theory consist in the agreement of predictions based on the theory with empirical evidence obtained either by experiment or by systematic observation. But what are the grounds which sanction the acceptance of mathematics? That is the question I propose to discuss in the present paper. For reasons which will become clear subsequently, I shall use the term "mathematics" here to refer to arithmetic, algebra, and analysis – to the exclusion, in particular, of geometry.[1]

2. Are the propositions of mathematics self-evident truths?

One of the several answers which have been given to our problem asserts that the truths of mathematics, in contradistinction to the hypotheses of empirical science, require neither factual evidence nor any other justification because they are "self-evident." This view, however, which ultimately relegates decisions as to mathematical truth to a feeling of self-evidence, encounters various difficulties. First of all, many mathematical theorems are so hard to establish that even to the specialist in the particular field they appear as anything but self-evident. Secondly, it is well known that some of the most interesting results of mathematics – especially in such fields as abstract set theory and topology – run counter to deeply ingrained intuitions and the customary kind of feeling of self-evidence. Thirdly, the existence of mathematical conjectures such as those of Goldbach and of Fermat, which are quite elementary in content and yet undecided up to this day, certainly shows that not all mathematical truths can be self-evident. And finally, even if self-evidence were

Reprinted from *The American Mathematical Monthly*, vol. 52 (1945), pp. 543–556, with the kind permission of the author and The Mathematical Association of America.

[1] A discussion of the status of geometry is given in Hempel 1945a: 7–17.

attributed only to the basic postulates of mathematics, from which all other mathematical propositions can be deduced, it would be pertinent to remark that judgments as to what may be considered as self-evident, are subjective; they may vary from person to person and certainly cannot constitute an adequate basis for decisions as to the objective validity of mathematical propositions.

3. Is mathematics the most general empirical science?

According to another view, advocated especially by John Stuart Mill, mathematics is itself an empirical science which differs from the other branches, such as astronomy, physics, chemistry, etc., mainly in two respects: its subject matter is more general than that of any other field of scientific research, and its propositions have been tested and confirmed to a greater extent than those of even the most firmly established sections of astronomy or physics. Indeed, according to this view, the degree to which the laws of mathematics have been borne out by the past experiences of mankind is so overwhelming that – unjustifiably – we have come to think of mathematical theorems as qualitatively different from the well confirmed hypotheses or theories of other branches of science: we consider them as certain, while other theories are thought of as at best "very probable" or very highly confirmed.

But this view, too, is open to serious objections. From a hypothesis which is empirical in character – such as, for example, Newton's law of gravitation – it is possible to derive predictions to the effect that under certain specified conditions certain specified observable phenomena will occur. The actual occurrence of these phenomena constitutes confirming evidence, their non-occurrence disconfirming evidence for the hypothesis. It follows in particular that an empirical hypothesis is theoretically disconfirmable; i.e., it is possible to indicate what kind of evidence, if actually encountered, would disconfirm the hypothesis. In the light of this remark, consider now a simple "hypothesis" from arithmetic: $3+2=5$. If this is actually an empirical generalization of past experiences, then it must be possible to state what kind of evidence would oblige us to concede the hypothesis was not generally true after all. If any disconfirming evidence for the given proposition can be thought of, the following illustration might well be typical of it: We place some microbes on a slide, putting down first three of them and then another two. Afterwards we count all the microbes to test whether in this instance 3 and 2 actually added up to 5. Suppose now that we counted 6 microbes altogether. Would we consider this as an empirical disconfirmation of the given proposition, or at least as a proof that it does not apply to

microbes? Clearly not; rather, we would assume we had made a mistake in counting or that one of the microbes had split in two between the first and the second count. But under no circumstances could the phenomenon just described invalidate the arithmetical proposition in question; for the latter asserts nothing whatever about the behavior of microbes; it merely states that any set consisting of $3+2$ objects may also be said to consist of 5 objects. And this is so because the symbols "$3+2$" and "5" denote the same number: they are synonymous by virtue of the fact that the symbols "2," "3," "5," and "+" are *defined* (or tacitly understood) in such a way that the above identity holds as a consequence of the meaning attached to the concepts involved in it.

4. The analytic character of mathematical propositions

The statement that $3+2=5$, then, is true for similar reasons as, say, the assertion that no sexagenarian is 45 years of age. Both are true simply by virtue of definitions or of similar stipulations which determine the meaning of the key terms involved. Statements of this kind share certain important characteristics: Their validation naturally requires no empirical evidence; they can be shown to be true by a mere analysis of the meaning attached to the terms which occur in them. In the language of logic, sentences of this kind are called analytic or true *a priori*, which is to indicate that their truth is logically independent of, or logically prior to, any experiential evidence.[2] And while the statements of empirical science, which are synthetic and can be validated only a posteriori, are constantly subject to revision in the light of new evidence, the truth of an analytic statement can be established definitely, once and for all. However, this characteristic "theoretical certainty" of analytic propositions has to be paid for at a high price: An analytic statement conveys no factual information. Our statement about sexagenarians, for example, asserts nothing that could possibly conflict with any factual evidence: it has no factual implications, no empirical content; and it is precisely for this reason that the statement can be validated without recourse to empirical evidence.

[2]The objection is sometimes raised that without certain types of experience, such as encountering several objects of the same kind, the integers and the arithmetical operations with them would never have been invented, and that therefore the propositions of arithmetic do have an empirical basis. This type of argument, however, involves a confusion of the logical and the psychological meaning of the term "basis." It may very well be the case that certain experiences occasion psychologically the formation of arithmetical ideas and in this sense form an empirical "basis" for them; but this point is entirely irrelevant for the logical questions as to the *grounds* on which the propositions of arithmetic may be accepted as true. The point made above is that no empirical "basis" or evidence whatever is needed to establish the truth of the propositions of arithmetic.

Let us illustrate this view of the nature of mathematical propositions by reference to another, frequently cited, example of a mathematical - or rather logical - truth, namely the proposition that whenever $a=b$ and $b=c$ then $a=c$. On what grounds can this so-called "transitivity of identity" be asserted? Is it of an empirical nature and hence at least theoretically disconfirmable by empirical evidence? Suppose, for example, that a, b, c are certain shades of green, and that as far as we can see, $a=b$ and $b=c$, but clearly $a \neq c$. This phenomenon actually occurs under certain conditions; do we consider it as disconfirming evidence for the proposition under consideration? Undoubtedly not; we would argue that if $a \neq c$, it is impossible that $a=b$ and also $b=c$; between the terms of at least one of these latter pairs, there must obtain a difference, though perhaps only a subliminal one. And we would dismiss the possibility of empirical disconfirmation, and indeed the idea that an empirical test should be relevant here, on the grounds that identity is a transitive relation by virtue of its definition or by virtue of the basic postulates governing it.[3] Hence the principle in question is true *a priori*.

5. Mathematics as an axiomatized deductive system

I have argued so far that the validity of mathematics rests neither on its alleged self-evidential character nor on any empirical basis, but derives from the stipulations which determine the meaning of the mathematical concepts, and that the propositions of mathematics are therefore essentially "true by definition." This latter statement, however, is obviously oversimplified and needs restatement and a more careful justification.

For the rigorous development of a mathematical theory proceeds not simply from a set of definitions but rather from a set of non-definitional propositions which are not proved within the theory; these are the postulates or axioms of the theory.[4] They are formulated in terms of certain basic or primitive concepts for which no definitions are provided within the theory. It is sometimes asserted that the postulates themselves represent "implicit definitions" of the primitive terms. Such a characterization of the postulates, however, is misleading. For while the postulates do limit, in a specific sense, the meanings that can possibly be ascribed to the primitives, any self-consistent postulate system admits, nevertheless, many different interpretations of the primitive terms (this will soon be illustrated), whereas a set of definitions in the strict sense of the word determines the meanings of the definienda in a unique fashion.

[3] A precise account of the definition and the essential characteristics of the identity relation may be found in Tarski 1941: chap. 3.

[4] For a lucid and concise account of the axiomatic method, see Tarski 1941: chap. 6.

Once the primitive terms and the postulates have been laid down, the entire theory is completely determined; it is derivable from its postulational basis in the following sense: Every term of the theory is definable in terms of the primitives, and every proposition of the theory is logically deducible from the postulates. To be entirely precise, it is necessary also to specify the principles of logic which are to be used in the proof of the propositions, i.e. in their deduction from the postulates. These principles can be stated quite explicitly. They fall into two groups: Primitive sentences, or postulates, of logic (such as: If p and q is the case, then p is the case), and rules of deduction or inference (including, for example, the familiar modus ponens rule and the rules of substitution which make it possible to infer, from a general proposition, any one of its substitution instances). A more detailed discussion of the structure and content of logic would, however, lead too far afield in the context of this article.

6. Peano's axiom system as a basis for mathematics

Let us now consider a postulate system from which the entire arithmetic of the natural numbers can be derived. This system was devised by the Italian mathematician and logician G. Peano (1858–1932). The primitives of this system are the terms "0," "number," and "successor." While, of course, no definition of these terms is given within the theory, the symbol "0" is intended to designate the number 0 in its usual meaning, while the term "number" is meant to refer to the natural numbers $0, 1, 2, 3 \cdots$ exclusively. By the successor of a natural number n, which will sometimes briefly be called n', is meant the natural number immediately following n in the natural order. Peano's system contains the following 5 postulates:

P1. 0 is a number.
P2. The successor of any number is a number.
P3. No two numbers have the same successor.
P4. 0 is not the successor of any number.
P5. If P is a property such that (a) 0 has the property P, and (b) whenever a number n has the property P, then the successor of n also has the property P, then every number has the property P.

The last postulate embodies the principle of mathematical induction and illustrates in a very obvious manner the enforcement of a mathematical "truth" by stipulation. The construction of elementary arithmetic on this basis begins with the definition of the various natural numbers. 1 is defined as the successor of 0, or briefly as $0'$; 2 as $1'$, 3 as $2'$, and so on. By virtue of P2, this process can be continued indefinitely; because of P3 (in combination with P5), it never leads back to one of the numbers previously defined, and in view of P4, it does not lead back to 0 either.

As the next step, we can set up a definition of addition which expresses in a precise form the idea that the addition of any natural number to some given number may be considered as a repeated addition of 1; the latter operation is readily expressible by means of the successor relation. This definition of addition runs as follows:

D1. (a) $n+0=n$; (b) $n+k'=(n+k)'$.

The two stipulations of this recursive definition completely determine the sum of any two integers. Consider, for example, the sum $3+2$. According to the definitions of the numbers 2 and 1, we have $3+2=3+1'=3+(0')'$; by D1 (b), $3+(0')'=(3+0')'=((3+0)')'$; but by D1 (a), and by the definitions of the numbers 4 and 5, $((3+0)')'=(3')'=4'=5$. This proof also renders more explicit and precise the comments made earlier in this paper on the truth of the proposition that $3+2=5$: Within the Peano system of arithmetic, its truth flows not merely from the definition of the concepts involved, but also from the postulates that govern these various concepts. (In our specific example, the postulates P1 and P2 are presupposed to guarantee that 1, 2, 3, 4, 5 are numbers in Peano's system; the general proof that D1 determines the sum of any two numbers also makes use of P5.) If we call the postulates and definitions of an axiomatized theory the "stipulations" concerning the concepts of that theory, then we may say now that the propositions of the arithmetic of the natural numbers are true by virtue of the stipulations which have been laid down initially for the arithmetical concepts. (Note, incidentally, that our proof of the formula "$3+2=5$" repeatedly made use of the transitivity of identity; the latter is accepted here as one of the rules of logic which may be used in the proof of any arithmetical theorem; it is, therefore, included among Peano's postulates no more than any other principle of logic.)

Now, the multiplication of natural numbers may be defined by means of the following recursive definition, which expresses in a rigorous form the idea that a product nk of two integers may be considered as the sum of k terms each of which equals n.

D2. (a) $n\cdot 0=0$; (b) $n\cdot k'=n\cdot k+n$.

It now is possible to prove the familiar general laws governing addition and multiplication, such as the commutative, associative, and distributive laws ($n+k=k+n$, $n\cdot k=k\cdot n$; $n+(k+l)=(n+k)+l$, $n\cdot(k\cdot l)=(n\cdot k)\cdot l$; $n\cdot(k+l)=(n\cdot k)+(n\cdot l)$). – In terms of addition and multiplication, the inverse operations of subtraction and division can then be defined. But it turns out that these "cannot always be performed"; i.e., in contradistinction to the sum and the product, the difference and the quotient are not defined for every couple of numbers; for example, $7-10$ and $7\div 10$ are undefined. This situation suggests an enlargement of the number system by the introduction of negative and of rational numbers.

It is sometimes held that in order to effect this enlargement, we have to "assume" or else to "postulate" the existence of the desired additional kinds of numbers with properties that make them fit to fill the gaps of subtraction and division. This method of simply postulating what we want has its advantages; but, as Bertrand Russell puts it, they are the same as the advantages of theft over honest toil (1919: 71); and it is a remarkable fact that the negative as well as the rational numbers can be obtained from Peano's primitives by the honest toil of constructing explicit definitions for them, without the introduction of any new postulates or assumptions whatsoever. Every positive and negative integer – in contradistinction to a natural number which has no sign – is definable as a certain set of ordered couples of natural numbers; thus, the integer $+2$ is definable as the set of all ordered couples (m,n) of natural numbers where $m=n+2$; the integer -2 is the set of all ordered couples (m,n) of natural numbers with $n=m+2$. – Similarly, rational numbers are defined as classes of ordered couples of integers. – The various arithmetical operations can then be defined with reference to these new types of numbers, and the validity of all the arithmetical laws governing these operations can be proved by virtue of nothing more than Peano's postulates and the definitions of the various arithmetical concepts involved.

The much broader system thus obtained is still incomplete in the sense that not every number in it has a square root, and more generally, not every algebraic equation whose coefficients are all numbers of the system has a solution in the system. This suggests further expansions of the number system by the introduction of real and finally of complex numbers. Again, this enormous extension can be effected by mere definition, without the introduction of a single new postulate.[5] On the basis thus obtained, the various arithmetical and algebraic operations can be defined for the numbers of the new system, the concepts of function, of limit, of derivative and integral can be introduced, and the familiar theorems pertaining to these concepts can be proved, so that finally the huge system of mathematics as here delimited rests on the narrow basis of Peano's system: Every concept of mathematics can be defined by

[5]For a more detailed account of the construction of the number system on Peano's basis, cf. Russell 1919: esp. chaps. 1 and 7 [chapters 1, 2, and 18 are reprinted in this volume. Eds.]. – A rigorous and concise presentation of that construction, beginning, however, with the set of all integers rather than that of the natural numbers, may be found in Birkhoff and MacLane 1941: chaps. 1, 2, 3, 5. – For a general survey of the construction of the number system, cf. also Young 1911: esp. lectures 10, 11, 12.

means of Peano's three primitives, and every proposition of mathematics can be deduced from the five postulates enriched by the definitions of the non-primitive terms.[6] These deductions can be carried out, in most cases, by means of nothing more than the principles of formal logic; the proof of some theorems concerning real numbers, however, requires one assumption which is not usually included among the latter. This is the so-called axiom of choice. It asserts that given a class of mutually exclusive classes, none of which is empty, there exists at least one class which has exactly one element in common with each of the given classes. By virtue of this principle and the rules of formal logic, the content of all of mathematics can thus be derived from Peano's modest system - a remarkable achievement in systematizing the content of mathematics and clarifying the foundations of its validity.

7. Interpretations of Peano's primitives

As a consequence of this result, the whole system of mathematics might be said to be true by virtue of mere definitions (namely, of the non-primitive mathematical terms) provided that the five Peano postulates are true. However, strictly speaking, we cannot, at this juncture, refer to the Peano postulates as propositions which are either true or false, for they contain three primitive terms which have not been assigned any specific meaning. All we can assert so far is that any specific interpretation of the primitives which satisfies the five postulates - i.e., turns them into true statements - will also satisfy all the theorems deduced from them. But for Peano's system, there are several - indeed, infinitely many - interpretations which will do this. For example, let us understand by 0 the origin of a half-line, by the successor of a point on that half-line the

[6] As a result of very deep-reaching investigations carried out by K. Gödel it is known that arithmetic, and *a fortiori* mathematics, is an incomplete theory in the following sense: While all those propositions which belong to the classical systems of arithmetic, algebra, and analysis can indeed be derived, in the sense characterized above, from the Peano postulates, there exist nevertheless other propositions which can be expressed in purely arithmetical terms, and which are true, but which cannot be derived from the Peano system. And more generally: For any postulate system of arithmetic (or of mathematics for that matter) which is not self-contradictory, there exist propositions which are true, and which can be stated in purely arithmetical terms, but which cannot be derived from that postulate system. In other words, it is impossible to construct a postulate system which is not self-contradictory, and which contains among its consequences all true propositions which can be formulated within the language of arithmetic.

This fact does not, however, affect the result outlined above, namely, that it is possible to deduce, from the Peano postulates and the additional definitions of non-primitive terms, all those propositions which constitute the classical theory of arithmetic, algebra, and analysis; and it is to these propositions that I refer above and subsequently as the propositions of mathematics.

point 1 cm. behind it, counting from the origin, and by a number any point which is either the origin or can be reached from it by a finite succession of steps each of which leads from one point to its successor. It can then readily be seen that all the Peano postulates as well as the ensuing theorems turn into true propositions, although the interpretation given to the primitives is certainly not the customary one, which was mentioned earlier. More generally, it can be shown that every progression of elements of any kind provides a true interpretation, or a "model," of the Peano system. This example illustrates our earlier observation that a postulate system cannot be regarded as a set of "implicit definitions" for the primitive terms: The Peano system permits of many different interpretations, whereas in everyday as well as in scientific language, we attach one specific meaning to the concepts of arithmetic. Thus, e.g., in scientific and in everyday discourse, the concept 2 is understood in such a way that from the statement "Mr. Brown as well as Mr. Cope, but no one else is in the office, and Mr. Brown is not the same person as Mr. Cope," the conclusion "Exactly two persons are in the office" may be validly inferred. But the stipulations laid down in Peano's system for the natural numbers, and for the number 2 in particular, do not enable us to draw this conclusion; they do not "implicitly determine" the customary meaning of the concept 2 or of the other arithmetical concepts. And the mathematician cannot acquiesce at this deficiency by arguing that he is not concerned with the customary meaning of the mathematical concepts; for in proving, say, that every positive real number has exactly two real square roots, he is himself using the concept 2 in its customary meaning, and his very theorem cannot be proved unless we presuppose more about the number 2 than is stipulated in the Peano system.

If therefore mathematics is to be a correct theory of the mathematical concepts in their intended meaning, it is not sufficient for its validation to have shown that the entire system is derivable from the Peano postulates plus suitable definitions; rather, we have to inquire further whether the Peano postulates are actually true when the primitives are understood in their customary meaning. This question, of course, can be answered only after the customary meaning of the terms "0," "natural number," and "successor" has been clearly defined. To this task we now turn.

8. Definition of the customary meaning of the concepts of arithmetic in purely logical terms

At first blush, it might seem a hopeless undertaking to try to define these basic arithmetical concepts without presupposing other terms of arithmetic, which would involve us in a circular procedure. However, quite rigorous definitions of the desired kind can indeed be formulated, and it can be shown that for the concepts so defined, all Peano postulates turn into true statements. This important result is due to the research of the German logician G. Frege (1848–1925) and to the subsequent systematic and detailed work of the contemporary English logicians and philosophers B. Russell and A. N. Whitehead. Let us consider briefly the basic ideas underlying these definitions.[7]

A natural number – or, in Peano's term, a number – in its customary meaning can be considered as a characteristic of certain *classes* of objects. Thus, e.g., the class of the apostles has the number 12, the class of the Dionne quintuplets the number 5, any couple the number 2, and so on. Let us now express precisely the meaning of the assertion that a certain class C has the number 2, or briefly, that $n(C)=2$. Brief reflection will show that the following definiens is adequate in the sense of the customary meaning of the concept 2: There is some object x and some object y such that (1) $x \in C$ (i.e., x is an element of C) and $y \in C$, (2) $x \neq y$, and (3) if z is any object such that $z \in C$, then either $z=x$ or $z=y$. (Note that on the basis of this definition it becomes indeed possible to infer the statement "The number of persons in the office is 2" from "Mr. Brown as well as Mr. Cope, but no one else is in the office, and Mr. Brown is not identical with Mr. Cope"; C is here the class of persons in the office.) Analogously, the meaning of the statement that $n(C)=1$ can be defined thus: There is some x such that $x \in C$, and any object y such that $y \in C$, is identical with x. Similarly, the customary meaning of the statement that $n(C)=0$ is this: There is no object such that $x \in C$.

The general pattern of these definitions clearly lends itself to the definition of any natural number. Let us note especially that in the definitions thus obtained, the definiens never contains any arithmetical term, but merely expressions taken from the field of formal logic, including the signs of identity and difference. So far, we have defined only the meaning of such phrases as "$n(C)=2$," but we have given no definition for the numbers $0, 1, 2, \ldots$ apart from this context. This desideratum can be met on the basis of the consideration that 2 is that property which is common to all couples, i.e., to all classes C such that $n(C)=2$. This common property may be conceptually represented by the class of all those classes which share this property. Thus we arrive at the definition: 2 is the class

[7]For a more detailed discussion, cf. Russell, 1919: chaps. 2, 3, 4. A complete technical development of the idea can be found in the great standard work in mathematical logic, Whitehead and Russell 1910–13. – For a very precise recent development of the theory, see Quine 1940. – A specific discussion of the Peano system and its interpretations from the viewpoint of semantics is included in Carnap 1939: esp. sections 14, 17, 18.

of all couples, i.e., the class of all classes C for which $n(C)=2$. – This definition is by no means circular because the concept of couple – in other words, the meaning of "$n(C)=2$" – has been previously defined without any reference to the number 2. Analogously, 1 is the class of all unit classes, i.e., the class of all classes C for which $n(C)=1$. Finally, 0 is the class of all null classes, i.e., the class of all classes without elements. And as there is only one such class, 0 is simply the class whose only element is the null class. Clearly, the customary meaning of any given natural number can be defined in this fashion.[8] In order to characterize the intended interpretation of Peano's primitives, we actually need, of all the definitions here referred to, only that of the number 0. It remains to define the terms "successor" and "integer."

The definition of "successor," whose precise formulation involves too many niceties to be stated here, is a careful expression of a simple idea which is illustrated by the following example: Consider the number 5, i.e., the class of all quintuplets. Let us select an arbitrary one of these quintuplets and add to it an object which is not yet one of its members. $5'$, the successor of 5, may then be defined as the number applying to the set thus obtained (which, of course, is a sextuplet). Finally, it is possible to formulate a definition of the customary meaning of the concept of natural number; this definition, which again cannot be given here, expresses, in a rigorous form, the idea that the class of the natural numbers consists of the number 0, its successor, the successor of that successor, and so on.

If the definitions here characterized are carefully written out – this is one of the cases where the techniques of symbolic, or mathematical, logic prove indispensable – it is seen that the definiens of every one of them contains exclusively terms from the field of pure logic. In fact, it is possible to state the customary interpretation of Peano's primitives, and thus also the meaning of every concept definable by means of them – and that includes every concept of mathematics – in terms of the following 7 expressions, in addition to variables such as "x" and "C": *not, and, if – then; for every object x* it is the case that ...; *there is some object x* such that ...; x is an *element* of class C; *the class of all things x such that....*

[8]The assertion that the definitions given above state the "customary" meaning of the arithmetical terms involved is to be understood in the logical, not the psychological sense of the term "meaning." It would obviously be absurd to claim that the above definitions express "what everybody has in mind" when talking about numbers and the various operations that can be performed with them. What is achieved by those definitions is rather a "logical reconstruction" of the concepts of arithmetic in the sense that if the definitions are accepted, then those statements in science and everyday discourse which involve arithmetical terms can be interpreted coherently and systematically in such a manner that they are capable of objective validation. The statement about the two persons in the office provides a very elementary illustration of what is meant here.

And it is even possible to reduce the number of logical concepts needed to a mere four: the first three of the concepts just mentioned are all definable in terms of "*neither–nor*," and the fifth is definable by means of the fourth and "*neither–nor*," thus, all the concepts of mathematics prove definable in terms of four concepts of pure logic. (The definition of one of the more complex concepts of mathematics in terms of the four primitives just mentioned may well fill hundreds or even thousands of pages; but clearly this affects in no way the theoretical importance of the result just obtained; it does, however, show the great convenience and indeed practical indispensability for mathematics of having a large system of highly complex defined concepts available.)

9. The truth of Peano's postulates in their customary interpretation

The definitions characterized in the preceding section may be said to render precise and explicit the customary meaning of the concepts of arithmetic. Moreover – and this is crucial for the question of the validity of mathematics – it can be shown that the Peano postulates all turn into true propositions if the primitives are construed in accordance with the definitions just considered.

Thus, P1 (0 is a number) is true because the class of all numbers – i.e., natural numbers – was defined as consisting of 0 and all its successors. The truth of P2 (The successor of any number is a number) follows from the same definition. This is true also of P5, the principle of mathematical induction. To prove this, however, we would have to resort to the precise definition of "integer" rather than the loose description given of that definition above. P4 (0 is not the successor of any number) is seen to be true as follows: By virtue of the definition of "successor," a number which is a successor of some number can apply only to classes which contain at least one element; but the number 0, by definition, applies to a class if and only if that class is empty. – While the truth of P1, P2, P4, P5 can be inferred from the above definitions simply by means of the principles of logic, the proof of P3 (No two numbers have the same successor) presents a certain difficulty. As was mentioned in the preceding section, the definition of the successor of a number n is based on the process of adding, to a class of n elements, one element not yet contained in that class. Now if there should exist only a finite number of things altogether then this process could not be continued indefinitely, and P3, which (in conjunction with P1 and P2) implies that the integers form an infinite set, would be false. Russell's way of meeting this difficulty was to introduce a special "axiom of infinity," which stipulates, in effect, the existence of infinitely many objects and thus makes P3 demonstrable (cf. Russell 1919: 24 and chap. 13). The axiom of infinity does not belong to the generally recognized laws of logic; but it is capable of expression in purely logical terms and may be treated as an additional postulate of logic.

10. Mathematics as a branch of logic

As was pointed out earlier, all the theorems of arithmetic, algebra, and analysis can be deduced from the Peano postulates and the definitions of those mathematical terms which are not primitives in Peano's system. This deduction requires only the principles of logic plus, in certain cases, the axiom of choice, which asserts that for any set of mutually exclusive non-empty sets $\alpha, \beta, \ldots$, there exists at least one set which contains exactly one element from each of the sets $\alpha, \beta, \ldots$, and which contains no other elements.[9] By combining this result with what has just been said about the Peano system, the following conclusion is obtained, which is also known as *the thesis of logicism concerning the nature of mathematics*:

Mathematics is a branch of logic. It can be derived from logic in the following sense:

a. all the concepts of mathematics, i.e. of arithmetic, algebra, and analysis, can be defined in terms of four concepts of pure logic.
b. All the theorems of mathematics can be deduced from those definitions by means of the principles of logic (including the axioms of infinity and choice).[10]

In this sense it can be said that the propositions of the system of mathematics as here delimited are true by virtue of the definitions of the mathematical concepts involved, or that they make explicit certain characteristics with which we have endowed our mathematical concepts by definition. The propositions of mathematics have, therefore, the same unquestionable certainty which is typical of such propositions as "All

[9]This only apparently self-evident postulate is used in proving certain theorems of set theory and of real and complex analysis; for a discussion of its significance and of its problematic aspects, see Russell 1919: chap. 12 (where it is called the multiplicative axiom), and Fraenkel 1919 (Dover): §16, sections 7 and 8.

[10]The principles of logic developed in Quine's work and in similar modern systems of formal logic embody certain restrictions as compared with those logical rules which had been rather generally accepted as sound until about the turn of the 20th century. At that time, the discovery of the famous paradoxes of logic, especially of Russell's paradox (cf. Russell 1919: chap. 13) revealed the fact that the logical principles implicit in customary mathematical reasoning involved contradictions and therefore had to be curtailed in one manner or another.

bachelors are unmarried," but they also share the complete lack of empirical content which is associated with that certainty: The propositions of mathematics are devoid of all factual content; they convey no information whatever on any empirical subject matter.

11. On the applicability of mathematics to empirical subject matter

This result seems to be irreconcilable with the fact that after all mathematics has proved to be eminently applicable to empirical subject matter, and that indeed the greater part of present-day scientific knowledge has been reached only through continual reliance on and application of the propositions of mathematics. – Let us try to clarify this apparent paradox by reference to some examples.

Suppose that we are examining a certain amount of some gas, whose volume v, at a certain fixed temperature, is found to be 9 cubic feet when the pressure p is 4 atmospheres. And let us assume further that the volume of the gas for the same temperature and $p=6$ at., is predicted by means of Boyle's law. Using elementary arithmetic we reason thus: For corresponding values of v and p, $vp=c$, and $v=9$ when $p=4$; hence $c=36$: Therefore, when $p=6$, then $v=6$. Suppose that this prediction is borne out by subsequent test. Does that show that the arithmetic used has a predictive power of its own, that its propositions have factual implications? Certainly not. All the predictive power here deployed, all the empirical content exhibited stems from the initial data and from Boyle's law, which asserts that $vp=c$ for *any* two corresponding values of v and p, hence also for $v=9$, $p=4$, and for $p=6$ and the corresponding value of v.[11] The function of the mathematics here applied is not predictive at all; rather, it is analytic or explicative: it renders explicit certain assumptions or assertions which are included in the content of the premises of the argument (in our case, these consist of Boyle's law plus the additional data); mathematical reasoning reveals that those premises contain – hidden in them, as it were, – an assertion about the case as yet unobserved. In accepting our premises – so arithmetic reveals – we have – knowingly or unknowingly – already accepted the implication that the p-value in question is 6. Mathematical as well as logical reasoning is a conceptual technique of making explicit what is implicitly contained in a set of premises. The conclusions to which this technique leads assert nothing that is *theoretically new* in the sense of not being contained in the content of the premises. But the results obtained may

[11] Note that we may say "hence" by virtue of the rule of substitution, which is one of the rules of logical inference.

well be *psychologically new*: we may not have been aware, before using the techniques of logic and mathematics, what we committed ourselves to in accepting a certain set of assumptions or assertions.

A similar analysis is possible in all other cases of applied mathematics, including those involving, say, the calculus. Consider, for example, the hypothesis that a certain object, moving in a specified electric field, will undergo a constant acceleration of 5 feet/sec^2. For the purpose of testing this hypothesis, we might derive from it, by means of two successive integrations, the prediction that if the object is at rest at the beginning of the motion, then the distance covered by it at any time t is $(5/2)t^2$ feet. This conclusion may clearly be psychologically new to a person not acquainted with the subject, but it is not theoretically new; the content of the conclusion is already contained in that of the hypothesis about the constant acceleration. And indeed, here as well as in the case of the compression of a gas, a failure of the prediction to come true would be considered as indicative of the factual incorrectness of at least one of the premises involved (*f.ex.*, of Boyle's law in its application to the particular gas), but never as a sign that the logical and mathematical principles involved might be unsound.

Thus, in the establishment of empirical knowledge, mathematics (as well as logic) has, so to speak, the function of a theoretical juice extractor: the techniques of mathematical and logical theory can produce no more juice of factual information than is contained in the assumptions to which they are applied; but they may produce a great deal more juice of this kind than might have been anticipated upon a first intuitive inspection of those assumptions which form the raw material for the extractor.

At this point, it may be well to consider briefly the status of those mathematical disciplines which are not outgrowths of arithmetic and thus of logic; these include in particular topology, geometry, and the various branches of abstract algebra, such as the theory of groups, lattices, fields, etc. Each of these disciplines can be developed as a purely deductive system on the basis of a suitable set of postulates. If P be the conjunction of the postulates for a given theory, then the proof of a proposition T of that theory consists in deducing T from P by means of the principles of formal logic. What is established by the proof is therefore not the truth of T, but rather the fact that T is true provided that the postulates are. But since both P and T contain certain primitive terms of the theory, to which no specific meaning is assigned, it is not strictly possible to speak of the truth of either P or T; it is therefore more adequate to state the point as follows: If proposition T is logically deduced from P, then every specific interpretation of the primitives which turns all the postulates of P into true statements, will also render T a true statement. – Up to this point, the analysis is exactly analogous to that of arithmetic as based on Peano's set of postulates. In the case of arithmetic, however, it proved possible to go a step further, namely to define the customary meanings of the primitives in terms of purely logical concepts and to show that the postulates – and therefore also the theorems – of arithmetic are unconditionally true by virtue of these definitions. An analogous procedure is not applicable to those disciplines which are not outgrowths of arithmetic: The primitives of the various branches of abstract algebra have no specific "customary meaning"; and if geometry in its customary interpretation is thought of as a theory of the structure of physical space, then its primitives have to be construed as referring to certain types of physical entities, and the question of the truth of a geometrical theory in this interpretation turns into an *empirical* problem.[12] For the purpose of applying any one of these nonarithmetical disciplines to some specific field of mathematics or empirical science, it is therefore necessary first to assign to the primitives some specific meaning and then to ascertain whether in this interpretation the postulates turn into true statements. If this is the case, then we can be sure that all the theorems are true statements too, because they are logically derived from the postulates and thus simply explicate the content of the latter in the given interpretation. – In their application to empirical subject matter, therefore, these mathematical theories no less than those which grow out of arithmetic and ultimately out of pure logic, have the function of an analytic tool, which brings to light the implications of a given set of assumptions but adds nothing to their content.

But while mathematics in no case contributes anything to the content of our knowledge of empirical matters, it is entirely indispensable as an instrument for the validation and even for the linguistic expression of such knowledge: the majority of the more far-reaching theories in empirical science – including those which lend themselves most eminently to prediction or to practical application – are stated with the help of mathematical concepts; the formulation of these theories makes use, in particular, of the number system, and of functional relationships among different metrical variables. Furthermore, the scientific test of these theories, the establishment of predictions by means of them, and finally their practical application, all require the deduction, from the general theory, of certain specific consequences; and such deduction would be entirely impossible without the techniques of mathematics which reveal

[12] For a more detailed discussion of this point, cf. Hempel 1945a.

what the given general theory implicitly asserts about a certain special case.

Thus, the analysis outlined on these pages exhibits the system of mathematics as a vast and ingenious conceptual structure without empirical content and yet an indispensable and powerful theoretical instrument for the scientific understanding and mastery of the world of our experience.

On the nature of mathematical reasoning

HENRI POINCARE

I

The very possibility of mathematical science seems an insoluble contradiction. If this science is only deductive in appearance, from whence is derived that perfect rigour which is challenged by none? If, on the contrary, all the propositions which it enunciates may be derived in order by the rules of formal logic, how is it that mathematics is not reduced to a gigantic tautology? The syllogism can teach us nothing essentially new, and if everything must spring from the principle of identity, then everything should be capable of being reduced to that principle. Are we then to admit that the enunciations of all the theorems with which so many volumes are filled are only indirect ways of saying that A is A?

No doubt we may refer back to axioms which are at the source of all these reasonings. If it is felt that they cannot be reduced to the principle of contradiction, if we decline to see in them any more than experimental facts which have no part or lot in mathematical necessity, there is still one resource left to us: we may class them among *a priori* synthetic views. But this is no solution of the difficulty - it is merely giving it a name; and even if the nature of the synthetic views had no longer for us any mystery, the contradiction would not have disappeared; it would have only been shirked. Syllogistic reasoning remains incapable of adding anything to the data that are given it; the data are reduced to axioms, and that is all we should find in the conclusions.

No theorem can be new unless a new axiom intervenes in its demonstration; reasoning can only give us immediately evident truths borrowed from direct intuition; it would only be an intermediary parasite. Should we not therefore have reason for asking if the syllogistic apparatus serves only to disguise what we have borrowed?

The contradiction will strike us the more if we open any book on mathematics; on every page the author announces his intention of generalizing some proposition already known. Does the mathematical

Excerpted and reprinted with the kind permission of the publisher from Henri Poincaré. *Science and Hypothesis* (New York: Dover Publications, Inc., 1952), pp. 1-19.

method proceed from the particular to the general, and, if so, how can it be called deductive?

Finally, if the science of number were merely analytical, or could be analytically derived from a few synthetic intuitions, it seems that a sufficiently powerful mind could with a single glance perceive all its truths; nay, one might even hope that some day a language would be invented simple enough for these truths to be made evident to any person of ordinary intelligence.

Even if these consequences are challenged, it must be granted that mathematical reasoning has of itself a kind of creative virtue, and is therefore to be distinguished from the syllogism. The difference must be profound. We shall not, for instance, find the key to the mystery in the frequent use of the rule by which the same uniform operation applied to two equal numbers will give identical results. All these modes of reasoning, whether or not reducible to the syllogism, properly so called, retain the analytical character, and *ipso facto*, lose their power.

II

The argument is an old one. Let us see how Leibnitz tried to show that two and two make four. I assume the number one to be defined, and also the operation $x+1$ - i.e., the adding of unity to a given number x. These definitions, whatever they may be, do not enter into the subsequent reasoning. I next define the numbers 2, 3, 4 by the equalities: -

$$(1)\quad 1+1=2; \qquad (2)\quad 2+1=3; \qquad (3)\quad 3+1=4,$$

and in the same way I define the operation $x+2$ by the relation;

$$(4)\quad x+2=(x+1)+1.$$

Given this, we have: -

$$2+2=(2+1)+1; \quad \text{(def. 4).}$$
$$(2+1)+1=3+1 \quad \text{(def. 2).}$$
$$3+1=4 \quad \text{(def. 3).}$$

whence

$$2+2=4 \qquad \text{Q.E.D.}$$

It cannot be denied that this reasoning is purely analytical. But if we ask a mathematician, he will reply: "This is not a demonstration properly so called; it is a verification." We have confined ourselves to bringing together one or other of two purely conventional definitions, and we have verified their identity; nothing new has been learned. *Verification* differs from proof precisely because it is analytical, and because it leads to nothing. It leads to nothing because the conclusion is nothing but the premisses translated into another language. A real proof, on the other hand, is fruitful, because the conclusion is in a sense more general than the premisses. The equality $2+2=4$ can be verified because it is particular. Each individual enunciation in mathematics may be always verified in the same way. But if mathematics could be reduced to a series of such verifications it would not be a science. A chess-player, for instance, does not create a science by winning a piece. There is no science but the science of the general. It may even be said that the object of the exact sciences is to dispense with these direct verifications.

III

Let us now see the geometer at work, and try to surprise some of his methods. The task is not without difficulty; it is not enough to open a book at random and to analyse any proof we may come across. First of all, geometry must be excluded, or the question becomes complicated by difficult problems relating to the rôle of the postulates, the nature and the origin of the idea of space. For analogous reasons we cannot avail ourselves of the infinitesimal calculus. We must seek mathematical thought where it has remained pure – i.e., in Arithmetic. But we still have to choose; in the higher parts of the theory of numbers the primitive mathematical ideas have already undergone so profound an elaboration that it becomes difficult to analyse them.

It is therefore at the beginning of Arithmetic that we must expect to find the explanation we seek; but it happens that it is precisely in the proofs of the most elementary theorems that the authors of classic treatises have displayed the least precision and rigour. We may not impute this to them as a crime; they have obeyed a necessity. Beginners are not prepared for real mathematical rigour; they would see in it nothing but empty, tedious subtleties. It would be waste of time to try to make them more exacting; they have to pass rapidly and without stopping over the road which was trodden slowly by the founders of the science.

Why is so long a preparation necessary to habituate oneself to this perfect rigour, which it would seem should naturally be imposed on all minds? This is a logical and psychological problem which is well worthy of study. But we shall not dwell on it; it is foreign to our subject. All I wish to insist on is, that we shall fail in our purpose unless we reconstruct the proofs of the elementary theorems, and give them, not the rough

form in which they are left so as not to weary the beginner, but the form which will satisfy the skilled geometer.

Definition of addition

I assume that the operation $x+1$ has been defined; it consists in adding the number 1 to a given number x. Whatever may be said of this definition, it does not enter into the subsequent reasoning.

We now have to define the operation $x+a$, which consists in adding the number a to any given number x. Suppose that we have defined the operation $x+(a-1)$; the operation $x+a$ will be defined by the equality: (1) $x+a=[x+(a-1)]+1$. We shall know what $x+a$ is when we know what $x+(a-1)$ is, and as I have assumed that to start with we know what $x+1$ is, we can define successively and "by recurrence" the operations $x+2$, $x+3$, etc. This definition deserves a moment's attention; it is of a particular nature which distinguishes it even at this stage from the purely logical definition; the equality (1), in fact, contains an infinite number of distinct definitions, each having only one meaning when we know the meaning of its predecessor.

Properties of addition

Associative. – I say that $a+(b+c)=(a+b)+c$; in fact, the theorem is true for $c=1$. It may then be written $a+(b+1)=(a+b)+1$; which, remembering the difference of notation, is nothing but the equality (1) by which I have just defined addition. Assume the theorem true for $c=\gamma$, I say that it will be true for $c=\gamma+1$. Let $(a+b)+\gamma=a+(b+\gamma)$, it follows that $[(a+b)+\gamma]+1=[a+(b+\gamma)]+1$; or by def. (1) – $(a+b)+(\gamma+1)=a+(b+\gamma+1)=a+[b+(\gamma+1)]$, which shows by a series of purely analytical deductions that the theorem is true for $\gamma+1$. Being true for $c=1$, we see that it is successively true for $c=2$, $c=3$, etc.

Commutative. – (1) I say that $a+1=1+a$. The theorem is evidently true for $a=1$; we can *verify* by purely analytical reasoning that if it is true for $a=\gamma$ it will be true for $a=\gamma+1$.[1] Now, it is true for $a=1$, and therefore is true for $a=2$, $a=3$, and so on. This is what is meant by saying that the proof is demonstrated "by recurrence."

(2) I say that $a+b=b+a$. The theorem has just been shown to hold good for $b=1$, and it may be verified analytically that if it is true for

[1] For $(\gamma+1)+1=(1+\gamma)+1=1+(\gamma+1)$. – [Tr.]

$b=\beta$, it will be true for $b=\beta+1$. The proposition is thus established by recurrence.

Definition of multiplication

We shall define multiplication by the equalities: (1) $a\times 1=a$. (2) $a\times b=[a\times(b-1)]+a$. Both of these include an infinite number of definitions; having defined $a\times 1$, it enables us to define in succession $a\times 2$, $a\times 3$, and so on.

Properties of multiplication

Distributive. – I say that $(a+b)\times c=(a\times c)+(b\times c)$. We can verify analytically that the theorem is true for $c=1$; then if it is true for $c=\gamma$, it will be true for $c=\gamma+1$. The proposition is then proved by recurrence.

Commutative. – (1) I say that $a\times 1=1\times a$. the theorem is obvious for $a=1$. We can verify analytically that if it is true for $a=a$, it will be true for $a=\alpha+1$.

(2) I say that $a\times b=b\times a$. The theorem has just been proved for $b=1$. We can verify analytically that if it be true for $b=\beta$ it will be true for $b=\beta+1$.

IV

This monotonous series of reasonings may now be laid aside; but their very monotony brings vividly to light the process, which is uniform, and is met again at every step. The process is proof by recurrence. We first show that a theorem is true for $n=1$; we then show that if it is true for $n-1$ it is true for n, and we conclude that it is true for all integers. We have now seen how it may be used for the proof of the rules of addition and multiplication – that is to say, for the rules of the algebraical calculus. This calculus is an instrument of transformation which lends itself to many more different combinations than the simple syllogism; but it is still a purely analytical instrument, and is incapable of teaching us anything new. If mathematics had no other instrument, it would immediately be arrested in its development; but it has recourse anew to the same process – i.e., to reasoning by recurrence, and it can continue its forward march. Then if we look carefully, we find this mode of reasoning at every step, either under the simple form which we have just given

to it, or under a more or less modified form. It is therefore mathematical reasoning *par excellence*, and we must examine it closer.

V

The essential characteristic of reasoning by recurrence is that it contains, condensed, so to speak, in a single formula, an infinite number of syllogisms. We shall see this more clearly if we enunciate the syllogisms one after another. They follow one another, if one may use the expression, in a cascade. The following are the hypothetical syllogisms: – The theorem is true of the number 1. Now, if it is true of 1, it is true of 2; therefore it is true of 2. Now, if it is true of 2, it is true of 3; hence it is true of 3, and so on. We see that the conclusion of each syllogism serves as the minor of its successor. Further, the majors of all our syllogisms may be reduced to a single form. If the theorem is true of $n-1$, it is true of n.

We see, then, that in reasoning by recurrence we confine ourselves to the enunciation of the minor of the first syllogism, and the general formula which contains as particular cases all the majors. This unending series of syllogisms is thus reduced to a phrase of a few lines.

It is now easy to understand why every particular consequence of a theorem may, as I have above explained, be verified by purely analytical processes. If, instead of proving that our theorem is true for all numbers, we only wish to show that it is true for the number 6 for instance, it will be enough to establish the first five syllogisms in our cascade. We shall require 9 if we wish to prove it for the number 10; for a greater number we shall require more still; but however great the number may be we shall always reach it, and the analytical verification will always be possible. But however far we went we should never reach the general theorem applicable to all numbers, which alone is the object of science. To reach it we should require an infinite number of syllogisms, and we should have to cross an abyss which the patience of the analyst, restricted to the resources of formal logic, will never succeed in crossing.

I asked at the outset why we cannot conceive of a mind powerful enough to see at a glance the whole body of mathematical truth. The answer is now easy. A chess-player can combine for four or five mores ahead; but, however extraordinary a player he may be, he cannot prepare for more than a finite number of moves. If he applies his faculties to Arithmetic, he cannot conceive its general truths by direct intuition alone; to prove even the smallest theorem he must use reasoning by recurrence, for that is the only instrument which enable us to pass from the finite to the infinite. This instrument is always useful, for it enables us to leap over as many stages as we wish; it frees us from the necessity of long, tedious, and monotonous verifications which would rapidly become impracticable. Then when we take in hand the general theorem it becomes indispensable, for otherwise we should ever by approaching the analytical verification without every actually reaching it. In this domain of Arithmetic we may think ourselves very far from the infinitesimal analysis, but the idea of mathematical infinity is already playing a preponderating part, and without it there would be no science at all, because there would be nothing general.

VI

The views upon which reasoning by recurrence is based may be exhibited in other forms; we may say, for instance, that in any finite collection of different integers there is always one which is smaller than any other. We may readily pass from one enunciation to another, and thus give ourselves the illusion of having proved that reasoning by recurrence is legitimate. But we shall always be brought to a full stop – we shall always come to an indemonstrable axiom, which will at bottom be but the proposition we had to prove translated into another language. We cannot therefore escape the conclusion that the rule of reasoning by recurrence is irreducible to the principle of contradiction. Nor can the rule come to us from experiment. Experiment may teach us that the rule is true for the first ten or the first hundred numbers, for instance; it will not bring us to the indefinite series of numbers, but only to a more or less long, but always limited, portion of the series.

Now, if that were all that is in question, the principle of contradiction would be sufficient, it would always enable us to develop as many syllogisms as we wished. It is only when it is a question of a single formula to embrace an infinite number of syllogisms that this principle breaks down, and there, too, experiment is powerless to aid. This rule, inaccessible to analytical proof and to experiment, is the exact type of the *a priori* synthetic intuition. On the other hand, we cannot see in it a convention as in the case of the postulates of geometry.

Why then is this view imposed upon us with such an irresistible weight of evidence? It is because it is only the affirmation of the power of the mind which knows it can conceive of the indefinite repetition of the same act, when the act is once possible. The mind has a direct intuition of this power, and experiment can only be for it an opportunity of using it, and thereby of becoming conscious of it.

But it will be said, if the legitimacy of reasoning by recurrence cannot be established by experiment alone, is it so with experiment aided by induction? We see successively that a theorem is true of the number 1, of

the number 2, of the number 3, and so on – the law is manifest, we say, and it is so on the same ground that every physical law is true which is based on a very large but limited number of observations.

It cannot escape our notice that here is a striking analogy with the usual processes of induction. But an essential difference exists. Induction applied to the physical sciences is always uncertain, because it is based on the belief in a general order of the universe, an order which is external to us. Mathematical induction – i.e., proof by recurrence – is, on the contrary, necessarily imposed on us, because it is only the affirmation of a property of the mind itself.

VII

Mathematicians, as I have said before, always endeavour to generalize the propositions they have obtained. To seek no further example, we have just shown the equality, $a+1=1+a$, and we then used it to establish the equality, $a+b=b+a$, which is obviously more general. Mathematics may, therefore, like the other sciences, proceed from the particular to the general. This is a fact which might otherwise have appeared incomprehensible to us at the beginning of this study, but which has no longer anything mysterious about it, since we have ascertained the analogies between proof by recurrence and ordinary induction.

No doubt mathematical recurrent reasoning and physical inductive reasoning are based on different foundations, but they move in parallel lines and in the same direction – namely, from the particular to the general.

Let us examine the case a little more closely. To prove the equality $a+2=2+a \ldots$ (1), we need only apply the rule $a+1=1+a$, twice, and write $a+2=a+1+1=1+a+1=1+1+a=2+a \ldots$ (2).

The equality thus deduced by purely analytical means is not, however, a simple particular case. It is something quite different. We may not therefore even say in the really analytical and deductive part of mathematical reasoning that we proceed from the general to the particular in the ordinary sense of the words. The two sides of the equality (2) are merely more complicated combinations than the two sides of the equality (1), and analysis only serves to separate the elements which enter into these combinations and to study their relations.

Mathematicians therefore proceed "by construction," they "construct" more complicated combinations. When they analyze these combinations, these aggregates, so to speak, into their primitive elements, they see the relations of the elements and deduce the relations of the aggregates themselves. The process is purely analytical, but it is not a passing from the general to the particular, for the aggregates obviously cannot be regarded as more particular than their elements.

Great importance has been rightly attached to this process of "construction," and some claim to see in it the necessary and sufficient condition of the progress of the exact sciences. Necessary, no doubt, but not sufficient! For a construction to be useful and not mere waste of mental effort, for it to serve as a stepping-stone to higher things, it must first of all possess a kind of unity enabling us to see something more than the juxtaposition of its elements. Or more accurately, there must be some advantage in considering the construction rather than the elements themselves. What can this advantage be? Why reason on a polygon, for instance, which is always decomposable into triangles, and not on elementary triangles? It is because there are properties of polygons of any number of sides, and they can be immediately applied to any particular kind of polygon. In most cases it is only after long efforts that those properties can be discovered, by directly studying the relations of elementary triangles. If the quadrilateral is anything more than the juxtaposition of two triangles, it is because it is of the polygon type.

A construction only becomes interesting when it can be placed side by side with other analogous constructions for forming species of the same genus. To do this we must necessarily go back from the particular to the general, ascending one or more steps. The analytical process "by construction" does not compel us to descend, but it leaves us at the same level. We can only ascend by mathematical induction, for from it alone we can learn something new. Without the aid of this induction, which in certain respects differs from, but is as fruitful as, physical induction, construction would be powerless to create science.

Let me observe, in conclusion, that this induction is only possible if the same operation can be repeated indefinitely. That is why the theory of chess can never become a science, for the different moves of the same piece are limited and do not resemble each other.

Mathematical truth

PAUL BENACERRAF

Although this symposium is entitled "Mathematical Truth," I will also discuss issues which are somewhat broader but which nevertheless have the notion of mathematical truth at their core, which themselves depend on how truth in mathematics is properly explained. The most important of these is mathematical knowledge. It is my contention that two quite distinct kinds of concerns have separately motivated accounts of the nature of mathematical truth: (1) the concern for having a homogeneous semantical theory in which semantics for the propositions of mathematics parallel the semantics for the rest of the language,[1] and (2) the concern that the account of mathematical truth mesh with a reasonable epistemology. It will be my general thesis that almost all accounts of the concept of mathematical truth can be identified with serving one or another of these masters *at the expense of the other.* Since I believe further that both concerns must be met by any adequate account, I find myself deeply dissatisfied with any package of semantics and epistemology that purports to account for truth and knowledge both within and outside of mathematics. For, as I will suggest, accounts of truth that treat mathematical and nonmathematical discourse in relevantly similar ways do so at the cost of leaving it unintelligible how we can have any mathematical knowledge whatsoever; whereas those which attribute to

Presented at a symposium on Mathematical Truth, sponsored jointly by the American Philosophical Association, Eastern Division, and the Association for Symbolic Logic, December 27, 1973. Various segments of an early (1967) version of this paper have been read at Berkeley, Harvard, Chicago Circle, Johns Hopkins, New York University, Princeton, and Yale. I am grateful for the help I received on these occasions, as well as for many comments from my colleagues at Princeton, both students and faculty. I am particularly indebted to Richard Grandy, Hartry Field, Adam Morton, and Mark Steiner. That these have not resulted in more significant improvements is due entirely to my own stubbornness. The present version is an attempt to summarize the essentials of the longer paper while making minor improvements along the way. The original version was written during 1967/68 with the generous support of the John Simon Guggenheim Foundation and Princeton University. This is gratefully acknowledged.

Reprinted with the kind permission of the editors from the *Journal of Philosophy* 70 (1973): 661–80.

[1] I am indulging here in the fiction that we *have* semantics for "the rest of language," or, more precisely, that the proponents of the views that take their impetus from this concern often think of themselves as having such semantics, at least for philosophically important segments of the language.

mathematical propositions the kinds of truth conditions we can clearly know to obtain, do so at the expense of failing to connect these conditions with any analysis of the sentences which shows how the assigned conditions are conditions of their *truth*. What this means must ultimately be spelled out in some detail if I am to make out my case, and I cannot hope to do that within this limited context. But I will try to make it sufficiently clear to permit you to judge whether or not there is likely to be anything in the claim.

I take it to be obvious that any philosophically satisfactory account of truth, reference, meaning, and knowledge must embrace them all and must be adequate for all the propositions to which these concepts apply.[2] An account of knowledge that *seems* to work for certain empirical propositions about medium-sized physical objects but which fails to account for more theoretical knowledge is unsatisfactory – not only because it is incomplete, but because it may be incorrect as well, even as an account of the things it seems to cover quite adequately. To think otherwise would be, among other things, to ignore the interdependence of our knowledge in different areas. And similarly for accounts of truth and reference. A theory of truth for the language we speak, argue in, theorize in, mathematize in, etc., should by the same token provide similar truth conditions for similar sentences. The truth conditions assigned to two sentences containing quantifiers should reflect in relevantly similar ways the contribution made by the quantifiers. Any departure from a theory thus homogeneous would have to be strongly motivated to be worth considering. Such a departure, for example, might manifest itself in a theory that gave an account of the contribution of quantifiers in mathematical reasoning different from that in normal everyday reasoning about pencils, elephants, and vice-presidents. David Hilbert urged such an account in "On the Infinite" [reprinted in this volume] which is discussed briefly below. Later on, I will try to say more about what conditions I would expect a satisfactory general theory of truth for our language to meet, as well as more about how such an account is to mesh with what I take to be a reasonable account of knowledge. Suffice it to say here that, although it

[2] I shall in fact have nothing to say about meaning in this paper. I believe that the concept is in much deserved disrepute, but I don't dismiss it for all that. Recent work, most notably by Kripke, suggests that what passed for a long time for meaning – namely the Fregean "sense" – has less to do with truth than Frege or his immediate followers thought it had. Reference is what is presumably most closely connected with truth, and it is for *this* reason that I will limit my attention to reference. If it is granted that change of reference can take place without a corresponding change in meaning, and that truth is a matter of reference, then talk of meaning is largely beside the point of the cluster of problems that concern us in this paper. These comments are not meant as arguments, but only as explanation.

will often be convenient to present my discussion in terms of theories of mathematical truth, we should always bear in mind that what is really at issue is our over-all philosophical view. I will argue that, *as an over-all view*, it is unsatisfactory – not so much because we lack a seemingly satisfactory account of mathematical truth or because we lack a seemingly satisfactory account of mathematical knowledge – as because we lack any account that satisfactorily brings the two together. I hope that it is possible ultimately to produce such an account; I hope further that this paper will help to bring one about by bringing into sharper focus some of the obstacles that stand in its way.

I. Two kinds of account

Consider the following two sentences:

(1) There are at least three large cities older than New York.
(2) There are at least three perfect numbers greater than 17.

Do they have the same logicogrammatical form? More specifically, are they both of the form

(3) There are at least three *FG*'s that bear *R* to *a*.

where 'There are at least three' is a numerical quantifier eliminable in the usual way in favor of existential quantifiers, variables, and identity; '*F*' and '*G*' are to be replaced by one-place predicates, '*R*' by a two-place predicate, and '*a*' by the name of an element of the universe of discourse of the quantifiers? What are the truth conditions of (1) and (2)? Are they relevantly parallel? Let us ignore both the vagueness of 'large' and 'older than' and the peculiarities of attributive-adjective constructions in English which make a large city not something large and a city but more (although not exactly) like something large *for* a city. With those complications set aside, it seems clear that (3) accurately reflects the form of (1) and thus that (1) will be true if and only if the thing named by the expression replacing '*a*' ('New York') bears the relation designated by the expression replacing '*R*' ('① is older than ②') to at least three elements (of the domain of discourse of the quantifiers) which satisfy the predicates replacing '*F*' and '*G*' ('large' and 'city', respectively). This, I gather, is what a suitable truth definition would tell us. And I think it's right. Thus, if (1) is true, it is because certain cities stand in a certain relation to each other, etc.

But what of (2)? May we use (3) in the same way as a matrix in spelling out the conditions of *its* truth? That sounds like a silly question to which the obvious answer is "Of course." Yet the history of the subject (the philosophy of mathematics) has seen many other answers. Some (including one of my past and present selves[3]), reluctant to face the consequences of combining what I shall dub such a "standard" semantical account with a platonistic view of the nature of numbers, have shied away from supposing that numerals are names and thus, by implication, that (2) is of the form (3). David Hilbert (1926) chose a different but equally divergent approach, in his case in an attempt to arrive at a satisfactory account of the use of the notion of infinity in mathematics. On one construal, Hilbert can be seen as segregating a class of statements and methods, those of "intuitive" mathematics, as those which needed no further justification. Let us suppose that these are all "finitely verifiable" in some sense that is not precisely specified. Statements of arithmetic that do not share this property – typically, certain statements containing quantifiers – are seen by Hilbert as instrumental devices for going from "real" or "finitely verifiable" statements to "real" statements, much as an instrumentalist regards theories in natural science as a way of going from observation sentences to observation sentences. These mathematically "theoretical" statements Hilbert called "ideal elements," likening their introduction to the introduction of points "at infinity" in projective geometry: they are introduced as a convenience to make simpler and more elegant the theory of the things you really care about. If their introduction does not lead to contradiction and if they have these other uses, then it is justified: hence the search for a consistency proof for the full system of first-order arithmetic.

If this is a reasonable, if sketchy, account of Hilbert's view, it indicates that he did not regard all quantified statements semantically on a par with one another. A semantics for arithmetic as he viewed it would be very hard to give. But hard or not, it would certainly not treat the quantifier in (2) in the same way as the quantifier in (1). Hilbert's view as outlined represents a flat denial that (3) is the model according to which (2) is constructed.

On other such accounts, the truth conditions for arithmetic sentences are given as their formal derivability from specified sets of axioms. When coupled with the desire to attribute a truth value to each closed sentence of arithmetic, these views were torpedoed by the incompleteness theorems. They could be restored at least to internal consistency either by the liberalization of what counts as derivability (e.g., by including the application of an ω-rule in permissible derivations) or by abandoning the

[3] See my "What Numbers Could Not Be," 1965 [reprinted in this volume].

desire for completeness. For lack of a better term and because they almost invariably key on the syntactic (combinatorial) features of sentences, I will call such views "combinatorial" views of the determinants of mathematical truth. The leading idea of combinatorial views is that of assigning truth values to arithmetic sentences on the basis of certain (usually proof-theoretic) syntactic facts about them. Often, truth is defined as (formal) derivability from certain axioms. (Frequently a more modest claim is made – the claim to truth-in-S, where S is the particular system in question.) In any event, in such cases truth is conspicuously not explained in terms of reference, denotation, or satisfaction. The "truth" predicate is syntactically defined.

Similarly, certain views of truth in arithmetic on which the Peano axioms are claimed to be "analytic" of the concept of number are also "combinatorial" in my sense. And so are conventionalist accounts, since what marks them as conventionalist is the contrast between them and the "realist" account that analyzes (2) by assimilating it to (1), via (3).

Finally, to make one further distinction, a view is not automatically "combinatorial" if it interprets mathematical propositions as being about combinatorial matters, either self-referentially or otherwise. For such a view might analyze mathematical propositions in a "standard" way in terms of the names and quantifiers they might contain and in terms of the properties they ascribe to the objects within their domains of discourse – which is to say that the underlying concept of *truth* is essentially Tarski's. The difference is that its proponents, although realists in their analysis of mathematical language, part ways with the platonists by construing the mathematical universe as consisting exclusively of mathematically unorthodox objects: Mathematics for them is limited to metamathematics, and that to syntax.

I will defer to later sections my assessment of the relative merits of these various approaches to the truth of such sentences as (2). At this point I wish only to introduce the distinction between, on the one hand, those views which attribute the obvious syntax (and the obvious semantics) to mathematical statements, and, on the other, those which, ignoring the apparent syntax and semantics, attempt to state truth conditions (or to specify and account for the existing distribution of truth values) on the basis of what are evidently non-semantic syntactic considerations. Ultimately I will argue that each kind of account has its merits and defects: each addresses itself to an important component of a coherent over-all philosophic account of truth and knowledge.

But what are these components, and how do they relate to one another?

II. Two conditions

A. The first component of such an over-all view is more directly concerned with the concept of truth. For present purposes we can state it as the requirement that there be an over-all theory of truth in terms of which it can be certified that the account of mathematical truth is indeed an account of mathematical *truth*. The account should imply truth conditions for mathematical propositions that are evidently conditions of their truth (and not *simply*, say, of their theoremhood in some formal system). This is not to *deny* that being a theorem of some system can be a truth condition for a given proposition or class of propositions. It is rather to require that any theory that proffers theoremhood as a condition of truth also *explain the connection between truth and theoremhood.*

Another way of putting this first requirement is to demand that any theory of mathematical truth be in conformity with a general theory of truth – a theory of truth theories, if you like – which certifies that the property of sentences that the account calls "truth" is indeed truth. This, it seems to me, can be done only on the basis of some general theory for at least the language as a whole (I assume that we skirt paradoxes in some suitable fashion). Perhaps the applicability of this requirement to the present case amounts only to a plea that the semantical apparatus of mathematics be seen as part and parcel of that of the natural language in which it is done, and thus that whatever *semantical* account we are inclined to give of names or, more generally, of singular terms, predicates, and quantifiers in the mother tongue include those parts of the mother tongue which we classify as mathematese.

I suggest that, if we are to meet this requirement, we shouldn't be satisfied with an account that fails to treat (1) and (2) in parallel fashion, on the model of (3). There may well be *differences*, but I expect these to emerge at the level of the analysis of the reference of the singular terms and predicates. I take it that we have only one such account: Tarski's, and that its essential feature is to define truth in terms of reference (or satisfaction) on the basis of a particular kind of syntactico-semantical analysis of the language, and thus that any putative analysis of mathematical truth must be an analysis of a concept which is a truth concept at least in Tarski's sense. Suitably elaborated, I believe this requirement to be inconsistent with all the accounts that I have termed "combinatorial." On the other hand, the account that assimilates (2) above to (1) and (3) obviously meets this condition, as do many variants of it.

B. My second condition on an over-all view presupposes that we have mathematical knowledge and that such knowledge is no less knowledge for being mathematical. Since our knowledge is of truths, or can be so construed, an account of mathematical truth, to be acceptable, must be consistent with the possibility of having mathematical knowledge: the conditions of the truth of mathematical propositions cannot make it impossible for us to know that they are satisfied. This is not to argue that there cannot be unknowable truths – only that not all truths can be unknowable, for we know some. The minimal requirement, then, is that a satisfactory account of mathematical truth must be consistent with the possibility that some such truths be knowable. To put it more strongly, the concept of mathematical truth, as explicated, must fit into an over-all account of knowledge in a way that makes it intelligible how we have the mathematical knowledge that we have. An acceptable semantics for mathematics must fit an acceptable epistemology. For example, if I know that Cleveland is between New York and Chicago, it is because there exists a certain relation between the truth conditions for that statement and my present "subjective" state of belief (whatever may be our accounts of truth and knowledge, they must connect with each other in this way). Similarly, in mathematics, it must be possible to link up what it is for p to be true with my belief that p. Though this is extremely vague, I think one can see how the second condition tends to rule out accounts that satisfy the first, and to admit many of those which do not. For a typical "standard" account (at least in the case of number theory or set theory) will depict truth conditions in terms of conditions on objects whose nature, as normally conceived, places them beyond the reach of the better understood means of human cognition (e.g., sense perception and the like). The "combinatorial" accounts, on the other hand, usually arise from a sensitivity to precisely this fact and are hence almost always motivated by epistemological concerns. Their virtue lies in providing an account of mathematical propositions based on the procedures we follow in justifying truth claims in mathematics: namely, proof. It is not surprising that *modulo* such accounts of mathematical truth, there is little mystery about how we can obtain mathematical knowledge. We need only account for our ability to produce and survey formal proofs.[4] However, squeezing the balloon at that point apparently makes it bulge on the side of truth: the more nicely we tie up the concept of proof, the more

[4]Properly done, this is of course an enormous task. Nevertheless it sets to one side accounting for the burden that is borne by the semantics of the system and by our understanding of it, concentrating instead on our ability to determine that certain formal objects have certain syntactically defined properties.

closely we link the definition of proof to combinatorial (rather than semantic) features, the more difficult it is to connect it up with the truth of what is being thus "proved" – or so it would appear.

These then are the two requirements. Separately, they seem innocuous enough. In the balance of this paper I will both defend them further and flesh out the argument that jointly they seem to rule out almost every account of mathematical truth that has been proposed. I will consider in turn the two basic approaches to mathematical truth that I mentioned above, weighing their relative advantages in light of the two fundamental principles that I am advancing. I hope that the principles themselves will receive some illumination and support as I do so.

III. The standard view

I call the "platonistic" account that analyzes (2) as being of the form (3) "the standard view." Its virtues are many, and it is worth enumerating them in some detail before passing to a consideration of its defects.

As I have already pointed out, this account assimilates the logical form of mathematical propositions to that of apparently similar empirical ones: empirical and mathematical propositions alike contain predicates, singular terms, quantifiers, etc.

But what of sentences that are not composed (or correctly analyzable as being composed of) names, predicates, and quantifiers? More directly to the point, what of sentences that do not belong to the kind of language for which Tarski has shown us how to define truth? I would say that we need for such languages (if there are any) an account of truth of the sort that Tarski supplied for "referential" languages. I assume that the truth conditions for the language (e.g., English) to which mathematese appears to belong are to be elaborated much along the lines that Tarski articulated. So, to some extent, the question posed in the previous section – How are truth conditions for (2) to be explained? – may be interpreted as asking whether the sublanguage of English in which mathematics is done is to receive the same sort of analysis as I am assuming is appropriate for much of the rest of English. If so, then the qualms I shall sketch in the next section concerning how to fit mathematical knowledge into an overall epistemology clearly apply – though they can perhaps be laid to rest by a suitable modification of theory. If, on the other hand, mathematese is not to be analyzed along referential lines, then we are clearly in need not only of an account of truth (i.e., a semantics) for this new kind of language, but also of a new *theory of truth theories* that relates truth for referential (quantificational) languages to truth for these new (newly

analyzed) languages. Given such an account, the task of accounting for mathematical knowledge would still remain; but it would presumably be an easier task, since the new semantical picture of mathematese would in most cases have been prompted by epistemological considerations. However, I do not give this alternative serious consideration in this paper because I don't think that anyone has ever actually chosen it. For to choose it is explicitly to consider *and reject* the "standard" interpretation of mathematical language, despite its superficial and initial plausibility, and then to provide an alternative semantics as a substitute.[5] The "combinatorial" theorists whom I discuss or refer to have usually wanted to have their cake and eat it too: they have not realized that the truth conditions that their account supplies for mathematical language have not been connected to the referential semantics which they assume is *also* appropriate for that language. Perhaps the closest candidate for an exception is Hilbert in the view I sketched briefly in the opening pages of this paper. But to pursue this further here would take us too far afield. Let us return, therefore, to our praise of the "standard view."

One of its primary advantages is that the truth definitions for individual mathematical theories thus construed will have the same recursion clauses as those employed for their less lofty empirical cousins. Or to put it another way, they can all be taken as parts of the same language for which we provide a single account for quantifiers regardless of the subdiscipline under consideration. Mathematical and empirical disciplines will not be distinguished in point of logical grammar. I have already underscored the importance of this advantage: it means that the logico-grammatical theory we employ in less recondite and more tractable domains will serve us well here. We can do with one, uniform, account and need not invent another for mathematics. This should hold true on virtually any grammatical theory coupled with semantics adequate to account for truth. My bias for what I call a Tarskian theory stems simply from the fact that he has given us the only viable systematic general account we have of truth. So, one consequence of the economy attending the standard view is that logical relations are subject to uniform treatment: they are invariant with subject matter. Indeed, they help define the concept of "subject matter." The same rules of inference may be used and their use accounted for by the same theory which provides us with our ordinary account of inference, thus avoiding a double standard. If we reject the standard view, mathematical inference will need a new and special account. As it is, standard uses of quantifier inferences are

[5] I sometimes think this is one of the things that Hilary Putnam wants to do in his stimulating article "Mathematics without Foundations," 1967a [reprinted in this volume].

justified by some sort of soundness proof. The formalization of theories in first-order logic requires for *its* justification the assurance (provided by the Completeness theorem) that all the logical consequences of the postulates will be forthcoming as theorems. The standard account delivers these guarantees. The obvious answers seem to work. To reject the standard view is to discard these answers. New ones would have to be found.

So much for the obvious virtues of this account. What are its faults?

As I suggested above, the principal defect of the standard account is that it appears to violate the requirement that our account of mathematical truth be susceptible to integration into our over-all account of knowledge. Quite obviously, to make out a persuasive case to this effect it would be necessary to sketch the epistemology I take to be at least roughly correct and on the basis of which mathematical truths, standardly construed, do not seem to constitute knowledge. This would require a lengthy detour through the general problems of epistemology. I will leave that to another time and content myself here with presenting a brief summary of the salient features of that view which bear most immediately on our problem.

IV. Knowledge

I favor a causal account of knowledge on which for *X* to know that *S* is true requires some causal relation to obtain between *X* and the referents of the names, predicates, and quantifiers of *S*. I believe in addition in a causal theory of *reference*, thus making the link to my saying knowingly that *S doubly* causal. I hope that what follows will dispel some of the fog which surrounds this formulation.

For Hermione to know that the black object she is holding is a truffle is for her (or at least requires her) to be in a certain (perhaps psychological) state.[6] It also requires the cooperation of the rest of the world, at least to the extent of permitting the object she is holding to be a truffle. Further – and this is the part I would emphasize – in the normal case, that the black object she is holding is a truffle must figure in a suitable way in a causal explanation of her belief that the black object she is holding is a truffle. But what is a "suitable way"? I will not try to say. A number of authors have published views that seem to point in this direction,[7] and, despite differences among them, there seems to be a core intuition which they share and which I think is correct although very difficult to pin down.

That some such view must be correct and underlies our conception of knowledge is indicated by what we would say under the following circumstances. It is claimed that *X* knows that *p*. We think that *X* could not know that *p*. What reasons can we offer in support of our view? If we are satisfied that *X* has normal inferential powers, that *p* is indeed true, etc., we are often thrown back on arguing that *X* could not have come into possession of the relevant evidence or reasons: that *X*'s four-dimensional space-time worm does not make the necessary (causal) contact with the grounds of the truth of the proposition for *X* to be in possession of evidence adequate to support the inference (if an inference was relevant). The proposition *p* places restrictions on what the world can be like. Our knowledge of the world, combined with our understanding of the restrictions placed by *p*, given by the truth conditions of *p*, will often tell us that a given individual could not have come into possession of evidence sufficient to come to know *p*, and we will thus deny his claim to the knowledge.

As an account of our knowledge about medium-sized objects, in the present, this is along the right lines. It will involve, causally, some direct reference to the facts known, and, through that, reference to these objects themselves. Furthermore, such knowledge (of houses, trees, truffles, dogs, and bread boxes) presents the clearest case and the easiest to deal with.

Other cases of knowledge can be explained as being based on inferences based on cases such as these, although there must evidently be interdependencies. This is meant to include our knowledge of general laws and theories, and, through them, our knowledge of the future and much of the past. This account follows closely the lines that have been proposed by empiricists, but with the crucial modification introduced by the explicitly causal condition mentioned above – but often left out of modern accounts, largely because of attempts to draw a careful distinction between "discovery" and "justification."

In brief, in conjunction with our other knowledge, we use *p* to determine the range of possible relevant evidence. We use what we know of *X* (the putative knower) to determine whether there could have been an appropriate kind of interaction, whether *X*'s current belief that *p* is

[6] If possible, I would like to avoid taking any stand on the cluster of issues in the philosophy of mind or psychology concerning the nature of psychological states. Any view on which Hermione can learn that the cat is on the mat by looking at a real cat on a real mat will do for my purposes. If looking at a cat on a mat puts Hermione into a state and you wish to call that state a physical, or psychological, or even physiological state, I will not object so long as it is understood that such a state, if it is her state of knowledge, is causally related in an appropriate way to the cat's having been on the mat when she looked. If there is no such state, then so much the worse for my view.

[7] To cite but a few: Harman 1973; Goldman 1967: 357–72; Skyrms 1967: 373–89.

causally related in a suitable way with what is the case because p is true – whether his evidence is drawn from the range determined by p. If not, then X could not know that p. The connection between what must be the case if p is true and the causes of X's belief can vary widely. But there is always *some* connection, and the connection relates the grounds of X's belief to the subject matter of p.

It must be possible to establish an appropriate sort of connection between the truth conditions of p (as given by an adequate truth definition for the language in which p is expressed) and the grounds on which p is said to be known, at least for propositions that one must *come to know* – that are not innate. In the absence of this, no connection has been established between *having those grounds* and *believing a proposition which is true*. Having those grounds cannot be fitted into an explanation of *knowing p*. The link between p and justifying a belief in *p on those grounds* cannot be made. But for that knowledge which is properly regarded as some form of justified true belief, then the link *must* be made. (Of course not *all* knowledge need be justified true belief for the point to be a sound one.)

It will come as no surprise that this has been a preamble to pointing out that combining *this* view of knowledge with the "standard" view of mathematical truth makes it difficult to see how mathematical knowledge is possible. If, for example, numbers are the kinds of entities they are normally taken to be, then the connection between the truth conditions for the statements of number theory and any relevant events connected with the people who are supposed to have mathematical knowledge cannot be made out.[8] It will be impossible to account for how anyone knows any properly number-theoretical propositions. This second condition on an account of mathematical truth will not be satisfied, because we have no account of how we know that the truth conditions for mathematical propositions obtain. One obvious answer – that some of these propositions are true if and only if they are derivable from certain axioms via certain rules – will not help here. For, to be sure, we can ascertain that *those* conditions obtain. But in such a case, what we lack is the link between truth and proof, when truth is directly defined in the standard way. In short, although it may be a truth condition of certain number-theoretic propositions that they be derivable from certain axioms according to certain rules, *that* this is a truth condition must also follow from the account of *truth* if the condition referred to is to help connect truth and knowledge, if it is by their proofs that we know mathematical truths.

[8]For an expression of healthy skepticism concerning this and related points, see Steiner 1973: 57–66.

Of course, given some set-theoretical account of arithmetic, both the syntax and the semantics of *arithmetic* can be set out so as superficially to meet the conditions we have laid down. But the regress that this invites is transparent, for the same questions must then be asked about the set theory in terms of which the answers are couched.

V. Two examples

There are many accounts of mathematical truth and mathematical knowledge. The theses I have been defending are intended to apply to them all. Rather than try to be comprehensive, however, I will devote these last few pages to the examination of two representative cases: one "standard" view and one "combinatorial" view. First the standard account, as expressed by one of its most explicit and lucid proponents, Kurt Gödel.

Gödel is thoroughly aware that on a realist (i.e., standard) account of mathematical truth our explanation of how we know the basic postulates must be suitably connected with how we interpret the referential apparatus of the theory. Thus, in discussing how we can resolve the continuum problem, once it has been shown to be undecidable by the accepted axioms, he paints the following picture:

> ...the objects of transfinite set theory...clearly do not belong to the physical world and even their indirect connection with physical experience is very loose...
>
> But, despite their remoteness from sense experience, we do have a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them and, moreover, to believe that a question not decidable now has meaning and may be decided in the future. [Gödel 1964; pp. 483–4 in this volume]

I find this picture both encouraging and troubling. What troubles me is that without an account of *how* the axioms "force themselves upon us as being true," the analogy with sense perception and physical science is without much content. For what is missing is *precisely* what my second principle demands: an account of the link between our cognitive faculties and the objects known. In physical science we have at least a start on such an account, and it is causal. We accept as knowledge only those beliefs which we can appropriately relate to our cognitive faculties. Quite appropriately, our conception of knowledge goes hand in hand with our conception of ourselves as knowers. To be sure, there is a *superficial* analogy. For, as Gödel points out, we "verify" axioms by deducing consequences from them concerning areas in which we seem to have more direct "perception" (clearer intuitions). But we are never told how we know even these, clearer, propositions. For example, the "verifiable" consequences of axioms of higher infinity are (otherwise undecidable) number-theoretical propositions which themselves are "verifiable" by computation up to any given integer. But the story, to be helpful anywhere, must tell us how we know statements of computational arithmetic – *if they mean what the standard account would have them mean.* And *that* we are not told. So the analogy is at best superficial.

So much for the troubling aspects. More important perhaps and what I find encouraging is the evident basic agreement which motivates Gödel's attempt to draw a parallel between mathematics and empirical science. He sees, I think, that something must be said to bridge the chasm, created by his realistic and platonistic interpretation of mathematical propositions, between the entities that form the subject matter of mathematics and the human knower. Instead of tinkering with the logical form of mathematical propositions or with the nature of the objects known, he postulates a special faculty through which we "interact" with these objects. We seem to agree on the analysis of the fundamental problem, but clearly disagree about the epistemological issue – about what avenues are open to us through which we may come to know things.

If our account of empirical knowledge is acceptable, it must be in part because it tries to make the connection evident in the case of our theoretical knowledge, where it is not *prima facie* clear how the causal account is to be filled in. Thus, when we come to mathematics, the absence of a coherent account of how our mathematical intuition is connected with the truth of mathematical propositions renders the over-all account unsatisfactory.

To introduce a speculative historical note, with some foundation in the texts, it might not be unreasonable to suppose that Plato had recourse to the concept of *anamnesis* at least in part to explain how, given the nature of the forms as he depicted them, one could ever have knowledge of them.[9]

The "combinatorial" view of mathematical truth has epistemological roots. It starts from the proposition that, whatever may be the "objects" of mathematics, our knowledge is obtained from proofs. Proofs are or can be (for some, must be) written down or spoken; mathematicians can survey them and come to agree that they *are* proofs. It is largely through these proofs that mathematical knowledge is obtained and transmitted.

[9] "The soul, then, as being immortal, and having been born again many times, and having seen all things that exist, whether in this world or in the world below, has knowledge of them all" (Plato, *Meno*, 81).

In short, this aspect of mathematical knowledge – its (essentially linguistic) means of production and transmission gives their impetus to the class of views that I call "combinatorial."

Noticing the role of proofs in the production of knowledge, it seeks the grounds of truth in the proofs themselves. Combinatorial views receive additional impetus from the realization that the platonist casts a shroud of mystery over how knowledge can be obtained at all. Add that realization to the belief that mathematics is a child of our own begetting (mathematical discovery, on these views, is seldom discovery about an independent reality), and it is not surprising that one looks for acts of conception to account for the birth. Many accounts of mathematical truth fall under this rubric. Perhaps almost all. I have mentioned several in passing, and I discussed Hilbert's view in "On the Infinite" very briefly. The final example I wish to consider is that of conventionalist accounts – the cluster of views that the truths of logic and mathematics are true (or can be made true) in virtue of explicit conventions where the conventions in question are usually the postulates of the theory. Once more, I will probably do them all an injustice by lumping together a number of views which their proponents would most certainly like to keep apart.

Quine, in his classic paper on this subject (1964, reprinted in this volume), has dealt clearly, convincingly, and decisively with the view that the truths of *logic* are to be accounted for as the products of convention – far better than I could hope to do here. He pointed out that, since we must account for infinitely many truths, the characterization of the eligible sentences as truths must be wholesale rather than retail. But wholesale characterization can proceed only via general principles – and, if we are supposed not to understand any logic at all, we cannot extract the individual instances from the general principles: we would need logic for such a task.

Persuasive as this may be, I wish to add another argument – not because I think this dead horse needs further flogging, but both because Quine's argument is limited to the case of logic and because the principal points I wish to bring out do not emerge sufficiently from it. Indeed, Quine grants the conventionalist certain principles I should like to deny him. In resting his case against conventionalism on the need for a wholesale characterization of infinitely many truths, Quine concedes that were there only finitely many truths to be reckoned with, the conventionalist might have a chance to make out his case. He says:

> If truth assignments could be made one by one, rather than an infinite number at a time, the above difficulty would disappear; truths of logic . . . would simply be

asserted severally by fiat, and the problem of inferring them from more general conventions would not arise. [p. 353 in this volume]

Thus, if some way could be found to make sentences of logic wear their truth values upon their sleeves, the objections to the conventionalist account of truth would disappear – for we would have determined truth values for all the sentences, which is all that one could ask.

I wonder, however, what such a sprinkling of the word 'true' would accomplish. Surely it cannot suffice in order to determine a concept of truth to assign values to each and every sentence of the language [suppose now that the language is set theory, in some first-order formalization] (let those with an even number of horseshoes be "true").

What would make such an assignment of the predicate 'true' the determination of *the concept of truth?* Simply the use of that monosyllable? Tarski has suggested that satisfaction of Convention T is a necessary and sufficient condition on a definition of truth for a particular language.[10] A mere (recursive) distribution of truth values can be parlayed into a truth theory that satisfies convention T. We can rest with that provided we are prepared to beg what I think is the main question and ignore the concept of translation that occurs in its (Convention T's) formulation. What would be missing, hard as it is to state, is the theoretical apparatus employed by Tarski in providing truth definitions, i.e., the analysis of truth in terms of the "referential" concepts of naming, predication, satisfaction, and quantification. A definition that does not proceed by the customary recursion clauses for the customary grammatical forms may not be adequate, even if it satisfies Convention T. The explanation must proceed through reference and satisfaction and, furthermore, must be supplemented with an account of reference itself. But the defense of this last claim is too involved a matter to take up here.[11]

The Quine of "Truth by Convention" felt that to determine the truth values of all the contexts that contain a word suffices to determine its reference. That *might* be so, if we already had the concept of truth and

[10]"The Concept of Truth in Formalized Languages," reprinted in Tarski 1956. Convention T is stated on pp. 187–8 as follows:

> CONVENTION T. A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called *an adequate definition of truth* if it has the following consequences:
>
> (α) all sentences which are obtained from the expression '$x \in \mathrm{Tr}$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage;
>
> (β) the sentence 'for any x, if $x \in \mathrm{Tr}$ then $x \in \mathrm{S}$' (in other words '$\mathrm{Tr} \subseteq \mathrm{S}$').

[11]For an excellent presentation of a similar view, see Field, 1972: 347–5.

chased the reference of the term that interested us down through the truth definition. But there seems to be something patently wrong with trying to fix the concept of truth *itself* in this way. In so doing, we throw away the very crutch which enables that method to work for other concepts. Truth and reference go hand in hand. Our concept of truth, insofar as we have one, proceeds through the mediation of the concepts Tarski has used to define it for the class of languages he has considered – the essence of Tarski's contribution goes much further than Convention T, but includes the schemata for the actual definition as well: an analysis of truth for a language that did not proceed through the familiar devices of predication, quantification, etc., should not give us satisfaction.

If this is at all near the mark, then it should be clear why "combinatorial" views of the nature of mathematical truth fail on my account. They avoid what seems to me to be the necessary route to an account of truth: through the subject matter of the propositions whose truth is being defined. Motivated by epistemological considerations, they come up with truth conditions whose satisfaction or nonsatisfaction mere mortals can ascertain; but the price they pay is their inability to connect these so-called "truth conditions" with the truth of the propositions for which they are conditions.

Even if it is granted that the truths of first-order logic do not stem from conventions, it might still be claimed that the rest of mathematics (set theory, for logicists; set theory, number theory, and other things for nonlogicists) consists of conventions formalized in first-order logic. This view too is subject to the objection that such a concept of convention need not bring *truth* along with it.[12] Indeed it is clear that it does not. For, even ignoring more general objections, once the logic is fixed, it becomes possible that the conventions thus stipulated turn out to be inconsistent. Hence it cannot be maintained that setting down conventions *guarantees* truth. But if it does not *guarantee* truth, what distinguishes those cases in which it provides for it from those in which it does not? Consistency cannot be the answer. To urge it as such is to *mis*construe the significance of the fact that *in*consistency is *proof* that truth has not been attained. The deeper reason once more is that postulational stipulation makes no connection between the propositions and their subject matter – stipulation does not provide for truth. At best, it limits the class of truth definitions (interpretations) consistent with the stipulations. But that is not enough.

[12]Identical arguments will apply to the view, perhaps indistinguishable from this one, that the postulates constitute *implicit definitions* of existing concepts (as opposed to stipulating how new ones are to be understood), if that is advanced to explain how we know the axioms to be true (we learned the language by learning *these* postulates).

To clarify the point, consider Russell's oft-cited dictum: "The method of 'postulating' what we want has many advantages; they are the same as the advantages of theft over honest toil" (Russell 1919: 71). On the view I am advancing, that's false. For with theft at least you come away with the loot, whereas implicit definition, conventional postulation, and their cousins are incapable of bringing truth. They are not only morally but practically deficient as well.

Models and reality

HILARY PUTNAM

In 1922 Skolem delivered an address before the Fifth Congress of Scandinavian Mathematics in which he pointed out what he called a "relativity of set-theoretic notions". This "relativity" has frequently been regarded as paradoxical; but today, although one hears the expression "the Löwenheim-Skolem Paradox", it seems to be thought of as only an *apparent* paradox, something the cognoscenti enjoy but are not seriously troubled by. Thus van Heijenoort writes, "The existence of such a 'relativity' is sometimes referred to as the Löwenheim-Skolem Paradox, but, of course, it is not a paradox in the sense of an antinomy; it is a novel and unexpected feature of formal systems." In this address I want to take up Skolem's arguments, not with the aim of refuting them but with the aim of extending them in somewhat the direction he seemed to be indicating. It is not my claim that the "Löwenheim-Skolem Paradox" is an antinomy *in formal logic;* but I shall argue that it *is* an antinomy, or something close to it, in *philosophy of language*. Moreover, I shall argue that the resolution of the antinomy – the only resolution that I myself can see as making sense – has profound implications for the great metaphysical dispute about realism which has always been the central dispute in the philosophy of language.

The structure of my argument will be as follows: I shall point out that in many different areas there are three main positions on reference and truth: there is the extreme Platonist position, which posits nonnatural mental powers of directly "grasping" forms (it is characteristic of this position that "understanding" or "grasping" is itself an irreducible and unexplicated notion); there is the verificationist position which replaces the classical notion of truth with the notion of verification or proof, at least when it comes to describing how the language is understood; and there is the moderate realist position which seeks to preserve the central-

Presidential Address delivered before the Winter Meeting of the Association for Symbolic Logic in Washington, D.C., December 29, 1977. I wish to thank bas van Fraassen for valuable comments on and criticisms of an earlier version.
Reprinted with the kind permission of the editors and the Association for Symbolic Logic, Inc., from *The Journal of Symbolic Logic* 45.3 (September 1980): 464–82.

ity of the classical notions of truth and reference without postulating nonnatural mental powers. I shall argue that it is, unfortunately, the *moderate* realist position which is put into deep trouble by the Löwenheim-Skolem Theorem and related model-theoretic results. Finally I will opt for verificationism as a way of preserving the outlook of scientific or empirical realism, which is totally jetisoned by Platonism, even though this means giving up *metaphysical* realism.

The Löwenheim-Skolem Theorem says that a satisfiable first-order theory (in a countable language) has a countable model. Consider the sentence:

(i) $-(\exists R)(R$ is *one-to-one. The domain of* $R \subset N$. *The range of values of* R *is* $S)$

where 'N' is a formal term for the set of all whole numbers and the three conjuncts in the matrix have the obvious first-order definitions.

Replace 'S' with the formal term for the set of all real numbers in your favorite formalized set theory. The (i) will be a *theorem* (proved by Cantor's celebrated "diagonal argument"). So your formalized set theory *says* that a certain set (call it "S") is nondenumerable. So S must *be* nondenumerable in all *models* of your set theory. So your set theory – say ZF (Zermelo-Fraenkel set theory) has only nondenumerable models. But this is impossible! For, by the Löwenheim-Skolem Theorem, *no* theory can have *only* nondenumerable models; if a theory has a nondenumerable model, it must have denumerably infinite ones as well. Contradiction.

The resolution of this apparent contradiction is not hard, as Skolem points out (and it is not this apparent contradiction that I referred to as an antinomy, or close to an antinomy). For (i) only "says" that S is nondenumerable when the quantifier $(\exists R)$ is interpreted as ranging over *all* relations on $N \times S$. But when we pick a *denumerable* model for the language of set theory, "$(\exists R)$" does not range over *all* relations; it ranges only over relations *in the model*. (i) only "says" that S is nondenumerable in a *relative* sense: the sense that the members of S cannot be put in one-to-one correspondence with a subset of N by any R *in the model*. A set S can be "nondenumerable" in this *relative* sense and yet be denumerable "in reality". This happens when there *are* one-to-one correspondences between S and N but all of them lie outside the given model. What is a "countable" set from the point of view of one model may be an uncountable set from the point of view of another model. As Skolem sums it up, "even the notions 'finite', 'infinite', 'simply infinite sequence' and so forth turn out to be merely relative within axiomatic set theory".

The philosophical problem. Up to a point all commentators agree on the significance of the existence of "unintended" interpretations, e.g., models in which what are "supposed to be" nondenumerable sets are "in reality" denumerable. All commentators agree that the existence of such models shows that the "intended" interpretation, or, as some prefer to speak, the "intuitive notion of a set", is not "captured" by the formal system. But if *axioms* cannot capture the "intuitive notion of a set", what possibly could?

A technical fact is of relevance here. The Löwenheim-Skolem Theorem has a strong form (the so-called "downward Löwenheim-Skolem Theorem"), which requires the axiom of choice to prove, and which tells us that a satisfiable first-order theory (in a countable language) has a countable model which is a submodel of any given model. In other words if we are given a nondenumerable model M for a theory, then we can find a countable model M' of that same theory in which the predicate symbols stand for the same relations (restricted to the smaller universe in the obvious way) as they did in the original model. The only difference between M and M' is that the "universe of M' – i.e., the totality that the variables of quantification range over – is a proper subset of the "universe" of M.

Now the argument that Skolem gave, and that shows that "the intuitive notion of a set" (if there is such a thing) is not "captured" by any formal system, shows that even a *formalization of total science* (if one could construct such a thing), or even a *formalization of all our beliefs* (whether they count as "science" or not), could not rule out denumerable interpretations, and, *a fortiori*, such a formalization could not rule out *unintended* interpretations of this notion.

This shows that "theoretical constraints", whether they come from set theory itself or from "total science", cannot fix the interpretation of the notion *set* in the "intended" way. What of "operational constraints"?

Even if we allow that there might be a *denumerable infinity* of measurable "magnitudes", and that each of them might be measured to *arbitrary rational accuracy* (which certainly seems a utopian assumption), it would not help. For, by the "downward Löwenheim-Skolem Theorem", we can find a countable submodel of the "standard" model (if there is such a thing) in which countably many predicates (each of which may have countably many things in its extension) have their extensions preserved. In particular, we can fix the values of countable many magnitudes at all rational space-time points, and still find a countable submodel which meets all the constraints. In short, there certainly seems to be a *countable* model of our *entire body of belief* which meets all operational constraints.

The philosophical problem appears at just this point. If we are told, "axiomatic set theory does not capture the intuitive notion of a set", then it is natural to think that *something else* – our "understanding" – does capture it. But what can our "understanding" come to, at least for a naturalistically minded philosopher, which is more than *the way we use our language?* The Skolem argument can be extended, as we have just seen, to show that the *total use of the language* (operational plus theoretical constraints) does not "fix" a unique "intended interpretation" any more than axiomatic set theory by itself does.

This observation can push a philosopher of mathematics in two different ways. If he is inclined to Platonism, he will take this as evidence that the mind has mysterious faculties of "grasping concepts" (or "perceiving mathematical objects") which the naturalistically minded philosopher will never succeed in giving an account of. But if he is inclined to some species of verificationism (i.e., to indentifying truth with verifiability, rather than with some classical "correspondence with reality") he will say, "Nonsense! All the 'paradox' shows is that our understanding of 'The real numbers are nondenumerable' consists in our knowing *what it is for this to be proved,* and not in our 'grasp' of a 'model'." In short, the extreme positions – Platonism and verificationism – seem to receive comfort from the Löwenheim-Skolem Paradox; it is only the "moderate" position (which tries to avoid mysterious "perceptions" of "mathematical objects" while retaining a classical notion of truth) which is in deep trouble.

An epistemological/logical digression. The problem just pointed out is a serious problem for any philosopher or philosophically minded logician who wishes to view set theory as the description of a determinate independently existing reality. But from a mathematical point of view, it may appear immaterial: what does it matter if there are many different models of set theory, and not a unique "intended model" *if they all satisfy the same sentences?* What we want to know as mathematicians is what sentences of set theory are true; we do not want to have the sets themselves in our hands.

Unfortunately, the argument can be extended. First of all, the theoretical constraints we have been speaking of must, on a naturalistic view, come from only two sources: they must come from something like human decision or convention, whatever the source of the "naturalness" of the decisions or conventions may be, or from human experience, both experience with nature (which is undoubtedly the source of our most basic "mathematical intuitions", even if it be unfashionable to say so), and experience with "doing mathematics". It is hard to believe that

either or both of these sources together can ever give us a *complete* set of axioms for set theory (since, for one thing, a complete set of axioms would have to be nonrecursive, and it is hard to envisage coming to have a nonrecursive set of axioms in the literature or in our heads even in the unlikely event that the human race went on forever doing set theory); and if a complete set of axioms is impossible, and the intended models (in the plural) are singled out only by theoretical plus operational constraints then sentences which are independent of the axioms which we will arrive at in the limit of set-theoretic inquiry really have *no* determinate truth value; they are just true in some intended models and false in others.

To show what bearing this fact may have on actual set-theoretic inquiry, I will have to digress for a moment into technical logic. In 1938 Gödel put forward a new axiom for set theory: the axiom "$V=L$". Here L is the class of all constructible sets, that is, the class of all sets which can be defined by a certain constructive procedure if we pretend to have names available for all the ordinals, however large. (Of course, this sense of "constructible" would be anathema to constructive mathematicians.) V is the universe of all sets. So "$V=L$" just says *all sets are constructible.* By considering the inner model for set theory in which "$V=L$" is true, Gödel was able to prove the relative consistency of ZF and ZF *plus* the axiom of choice and the generalized continuum hypothesis.

"$V=L$" is certainly an important sentence, mathematically speaking. Is it *true?*

Gödel briefly considered proposing that we *add* "$V=L$" to the accepted axioms for set theory, as a sort of meaning stipulation, but he soon changed his mind. His later view was that "$V=L$" is *really* false, even though it is consistent with set theory, if set theory is itself consistent.

Gödel's intuition is widely shared among working set theorists. But does this "intuition" make sense?

Let *MAG* be a countable set of physical magnitudes which includes all magnitudes that sentient beings in this physical universe can actually measure (it certainly seems plausible that we cannot hope to measure more than a countable number of physical magnitudes). Let *OP* be the "correct" assignment of values; that is, the assignment which assigns to each member of *MAG* the value that that magnitude actually has at each rational space-time point. Then all the information "operational constraints" might give us (and, in fact, infinitely more) is coded into *OP*.

One technical term: an *ω-model* for a set theory is a model in which the *natural numbers* are ordered as they are "supposed to be"; that is, the sequence of "natural numbers" of the model is an ω-sequence.

Now for a small theorem.[1]

THEOREM. *ZF plus $V=L$ has an ω-model which contains any given countable set of real numbers.*

PROOF. Since a countable set of reals can be coded as a single real by well-known techniques, it suffices to prove that *for every real s, there is an M such that M is an ω-model for ZF plus $V=L$ and s is represented in M.*

By the "downward Löwenheim-Skolem Theorem", this statement is true if and only if the following statement is:

For every real s, there is a countable M such that M is an ω-model for ZF plus $V=L$ and s is represented in M.

Countable structures with the property that the "natural numbers" of the structure form an ω-sequence can be coded as reals by standard techniques. When this is properly done, the predicate "*M is an ω-model for ZF plus $V=L$ and s is represented in M*" becomes a two-place *arithmetical* predicate of *reals M, s*. The above sentence thus has the logical form (*for every real s*) (*there is a real M*) $(\cdots M, s, \cdots)$. In short, the sentence is a Π_2-sentence.

Now, consider this sentence *in the inner model $V=L$*. For every *s in the inner model* – that is, for every s in L – there is a model – namely L itself – which satisfies "$V=L$" and contains s. By the downward Löwenheim-Skolem Theorem, there is a countable submodel which is elementary equivalent to L and contains s. (Strictly speaking, we need here not just the downward Löwenheim-Skolem Theorem, but the "Skolem hull" construction which is used to prove that theorem.) By Gödel's work, this countable submodel itself lies in L, and as is easily verified, so does the real that codes it. So the above Π_2-sentences is true in the inner model $V=L$.

But Schoenfield has proved that Π_2-sentences are *absolute:* if a Π_2-sentence is true in L, then it must be true in V. So the above sentence is true in V. □

What makes this theorem startling is the following reflection: suppose that Gödel is right, and "$V=L$" is *false* ("in reality"). Suppose that there is, in fact, a *non-constructible real number* (as Gödel also believes). Since the predicate "is constructible" is absolute in *β-models* – that is, in

[1] Barwise (1971) has proved the much stronger theorem that every countable model of ZF has a proper end extension which is a model of $ZF+V=L$. The theorem in the text was proved by me before 1963.

models in which the "wellorderings" *relative to the model* are wellorderings "in reality" (recall Skolem's "relativity of set-theoretic notions"!), no model containing such a nonconstructible *s* can satisfy "*s* is constructible" and be a *β-model*. But, by the above theorem, a model containing *s can* satisfy "*s* is constructible" (because it satisfies "$V=L$", and "$V=L$" says *everything* is constructible) and be an *ω-model*.

Now, suppose we formalize *the entire language of science* within the set theory ZF *plus $V=L$*. Any model for ZF which contains an abstract set isomorphic to *OP* can be extended to a model for this formalized language of science which is *standard with respect to OP* – hence, even if *OP* is nonconstructible "in reality", we can find a model *for the entire language of science* which satisfies *everything is constructible* and which assigns the correct values to all the physical magnitudes in *MAG* at all rational space-time points.

The claim Gödel makes is that "$V=L$" is false "in reality". But what on earth can this mean? It must mean, at the very least, that in the case just envisaged, the model we have described in which "$V=L$" holds would not be *the intended model*. But why not? It satisfies all theoretical constraints; and we have gone to great length to make sure it satisfies all operational constraints as well.

Perhaps someone will say that "$V\neq L$" (or something which implies that V does not equal L) should be added to the axioms of ZF as an additional "theoretical constraint". (Gödel often speaks of new axioms someday becoming evident.) But, while this may be acceptable from a nonrealist standpoint, it can hardly be acceptable from a realist standpoint. For the realist standpoint is that there is *a fact of the matter* – a fact independent of our legislation – as to whether $V=L$ or not. A realist like Gödel holds that we have access to an "intended interpretation" of ZF, where the access is not simply by linguistic stipulation.

What the above argument shows is that if the "intended interpretation" is fixed only by theoretical plus operational constraints, then if "$V\neq L$" does not follow from those constraints – if we do not *decide* to make $V=L$ true or to make $V=L$ false – then there will be "intended" models in which $V=L$ is *true*. If I am right, then the "relativity of set-theoretic notions" extends to a *relativity of the truth value of* "$V=L$" (and, by similar arguments, of the axiom of choice and the continuum hypothesis as well).

Operational constraints and counterfactuals. It may seem to some that there is a major equivocation in the notion of what *can* be measured, or observed, which endangers the apparently crucial claim that the evidence we *could* have amounts to at most denumerably many facts. Imagine a measuring apparatus that simply detects the presence of a particle within a finite volume dv around its own geometric center during each full minute on its clock. Certainly it comes up with at most denumerably many reports (each *yes* or *no*) even if it is left to run forever. But how many are the facts it *could* report? Well, if it were jiggled a little, by chance let us say, its geometric center would shift r centimeters in a given direction. It would then report totally different facts. Since for each number r it could be jiggled that way, the number of reports it could produce is nondenumerable – and it does not matter to this that we, and the apparatus itself, are incapable of distinguishing every real number r from every other one. The problem is simply one of scope for the modal word "can". In my argument, I must be identifying what I call observational constraints, not with the totality of facts that could be registered by observation – i.e., ones that either will be registered, or would be registered if certain chance perturbations occurred – but with the totality of facts that will in actuality be registered or observed, whatever those be.

In reply, I would point out that even if the measuring apparatus *were* jiggled r centimeters in a given direction, we could only know the real number r to some rational approximation. Now, if the intervals involved are all rational, there are only *countably* many facts of the form: *if action A* (an action described with respect to place, time, and character up to some finite "tolerance") were performed, then the result $r \pm \epsilon$ (a result described up to some rational tolerance) *would be obtained with probability in the interval a, b*. To know all facts of this form would be to know the *probability distribution* of all possible observable results of all possible actions. Our argument shows that a model could be constructed which agrees with all of these facts.

There is a deeper point to be made about this objection, however. Suppose we "first orderize" counterfactual talk, say, by including *events* in the ontology of our theory and introducing a predicate ("subjunctively necessitates") for the counter-factual connection between unactualized event types at a given place-time. Then our argument shows that a model exists which fits all the facts that will actually be registered or observed and fits our theoretical constraints, and this model *induces* an interpretation of the counterfactual idiom (a "similarity metric on possible worlds", in David Lewis' theory) which renders true just the counterfactuals that are true according to some completion of our theory. Thus appeal to counterfactual observations cannot rule out any models at all unless the interpretation of the counterfactual idiom itself is *already* fixed by something beyond operational and theoretical constraints.

(A related point is made by Wittgenstein in his *Philosophical Investigations:* talk about what an ideal machine – or God – could compute is talk *within* mathematics – in disguise – and cannot serve to fix the interpretation of mathematics. "God", too, has many interpretations.)

"Decision" and "convention". I have used the word "decision" in connection with open questions in set theory, and obviously this is a poor word. One cannot simply sit down in one's study and "decide" that "$V = L$" is to be true, or that the axiom of choice is to be true. Nor would it be appropriate for the mathematical community to call an international convention and legislate these matters. Yet, it seems to me that if we encountered an extra-terrestrial species of intelligent beings who had developed a high level of mathematics, and it turned out that they *rejected* the axiom of choice (perhaps because of the Tarski-Banach Theorem[2]), it would be wrong to regard them as simply making a *mistake*. To do *that* would, on my view, amount to saying that acceptance of the axiom of choice is built into our notion of rationality itself; that does not seem to me to be the case. To be sure, our acceptance of choice is not arbitrary; all kinds of "intuitions" (based, most likely, on experience with the finite) support it; its mathematical fertility supports it; but none of this is *so* strong that we could say that an equally successful culture which based *its* mathematics on principles *incompatible* with choice (e.g., on the so-called axiom of determinacy[4]) was *irrational*.

But if both systems of set theory – ours and the extra-terrestrials' – count as *rational*, what sense does it make to call one *true* and the others *false?* From the Platonist's point of view there is no trouble in answering this question. "The axiom of choice is true – true in *the* model", he will say (if he believes the axiom of choice). "We are right and the extra-terrestrials are wrong." But what is *the* model? If the intended model is singled out by theoretical and operational constraints, then, first, "the" intended model is plural not singular (so the "the" is inappropriate – our theoretical and operational constraints fit many models, not just one,

[2]This is a very counterintuitive consequence of the axiom of choice. Call two objects A, B "congruent by finite decomposition" if they can be divided into finitely many disjoint point sets $A_1, \ldots, A_n$, $B_1, \ldots B_n$, such that $A = A_1 \cup A_2 \cup \cdots \cup A_n$, $B = B_1 \cup B_2 \cup \cdots \cup B_n$, and (for $i = 1, 2, \ldots, n$) A_i is congruent to B_i. Then Tarski and Banach showed that *all spheres are congruent by finite decomposition.*

[3]This axiom, first studied by J. Mycielski (1964), asserts that infinite games with perfect information are determined, i.e. there is a winning strategy for either the first or second player. AD (the axiom of determinacy) implies the existence of a nontrivial countably additive two-valued measure on the real numbers, contradicting a well-known consequence of the axiom of choice.

and so do those of the extra-terrestrials as we saw before). Secondly, the intended models for us do satisfy the axiom of choice and the extra-terrestrially intended models do not; we are not talking about the same models, so there is no question of a "mistake" on one side or the other.

The Platonist will reply that what this really shows is that we have some mysterious faculty of "grasping concepts" (or "intuiting mathematical objects") and it is *this* that enables us to fix a model as *the* model, and not just operational and theoretical constraints; but this appeal to mysterious faculties seems both unhelpful as epistemology and unpersuasive as science. What neural process, after all, could be described as the perception of a mathematical object? Why of *one* mathematical object rather than another? I do not doubt that *some* mathematical axioms are built in to our notion of rationality ("every number has a successor"); but, if the axiom of choice and the continuum hypothesis are not, then, I am suggesting, Skolem's argument, or the foregoing extension of it, casts doubt on the view that these statements have a truth value independent of the theory in which they are embedded.

Now, suppose this is right and the axiom of choice is true when taken in the sense that it receives from *our* embedding theory and false when taken in the sense that it receives from extra-terrestrial theory. Urging this relativism is not advocating *unbridled* relativism; I do not doubt that there are some objective (if evolving) canons of rationality; I simply doubt that we would regard them as settling this sort of question, let alone as singling out *one* unique "rationally acceptable set theory". If this is right, then one is inclined to say that the extra-terrestrials have decided to let the axiom of choice be false and we have decided to let it be true; or that we have different "conventions"; but, of course, none of these words is literally right. It may well be the case that the idea that statements have their truth values *independent* of embedding theory is so deeply built into our ways of talking that there is simply no "ordinary language" word or short phrase which refers to the theory-dependence of meaning and truth. Perhaps this is why Poincaré was driven to exclaim "Convention, yes! Arbitrary, no!" when he was trying to express a similar idea in another context.

Is the problem a problem with the notion of a "set"? It would be natural to suppose that the problem Skolem points out, the problem of a surprising "relativity" of our notions, has to do with the notion of a "set", given the various problems which are *known* to surround *that* notion, or, at least, has to do with the problem of reference to "mathematical objects". But this is not so.

To see why it is not so, let us consider briefly the vexed problem of reference to theoretical entities in physical science. Although this may seem to be a problem more for philosophers of science or philosophers of language than for logicians, it is a problem whose logical aspects have frequently been of interest to logicians, as is witnessed by the expressions "Ramsey sentence", "Craig translation", etc. Here again, the realist – or, at least, the hard-core metaphysical realist – wishes it to be the case that *truth* and *rational acceptability* should be *independent* notions. He wishes it to be the case that what, e.g., electrons *are* should be distinct (and possibly different from) what we believe them to be or even what we would believe them to be given the best experiments and the epistemically best theory. Once again, the realist – the hard-core metaphysical realist – holds that our intentions single out "the" model, and that our beliefs are then either true or false in "the" model *whether we can find out their truth values or not.*

To see the bearing of the Löwenheim-Skolem Theorem (or of the intimately related Gödel Completeness Theorem and its model-theoretic generalizations) on this problem, let us again do a bit of model construction. This time the operational constraints have to be handled a little more delicately, since we have need to distinguish operational concepts (concepts that describe what we see, feel, hear, etc., as we perform various experiments, and also concepts that describe our acts of picking up, pushing, pulling, twisting, looking at, sniffing, listening to, etc.) from nonoperational concepts.

To describe our operational constraints we shall need three things. First, we shall have to fix a sufficiently large "observational vocabulary". Like the "observational vocabulary" of the logical empiricists, we will want to include in this set – call it the set of "O-terms" – such words as "red", "touches", "hard", "push", "look at", etc. Second, we shall assume that there *exists* (whether we can define it or not) a set of *S* which can be taken to be the set of macroscopically observable things and events (observable with the human sensorium, that means). The notion of an observable thing or event is surely vague; so we shall want *S* to be a generous set, that is, God is to err in the direction of counting too many things and events as "observable for humans" when He defines the set *S*, if it is necessary to err in either direction, rather than to err in the direction of leaving out some things that might be counted as borderline "observables". If one is a realist, then such a set *S* must exist, of course, even if our knowledge of the world and the human sensorium does not permit *us* to define it at the present time. The reason we allow *S* to contain events (and not just things) is that, as Richard Boyd has pointed out, some of the entities we can directly observe are *forces* – we can *feel* forces – and forces are not objects. But I assume that forces can be construed as predicates of either objects, e.g., our bodies, or of suitable events.

The third thing we shall assume given is a valuation (call it, once again '*OP*') which assigns the correct truth value to each n-place O-term (for $n=1,2,3,\ldots$) on each n-tuple of elements of S on which it is defined. O-terms are in general also defined on things not in S; for example, two molecules too small to see with the naked eye may touch, a dust-mote too small to see may be black, etc. Thus *OP* is a *partial* valuation in a double sense; it is defined on only a subset of the predicates of the language, namely the O-terms, and even on these it only fixes a part of the extension, namely the extension of $T \restriction S$ (the restriction of T to S), for each O-term T.

Once again, it is the valuation *OP* that captures our "operational constraints". Indeed, it captures these "from above", since it may well contain *more* information than we could actually get by using our bodies and our senses in the world.

What shall we do about "theoretical constraints"? Let us assume that there exists a possible formalization of present-day total science, call it 'T', and also that there exists a possible formalization of *ideal* scientific theory, call it 'T_I'. T_I is to be "ideal" in the sense of being *epistemically* ideal *for humans*. Ideality, in this sense, is a rather vague notion; but we shall assume that, when God makes up T_I, He constructs a theory which it would be rational for scientists to accept, or which is a limit of theories that it would be rational to accept, as more and more evidence accumulates, and also that he makes up a theory which is compatible with the valuation *OP*.

Now, the theory T is, we may suppose, well confirmed at the present time, and hence rationally acceptable on the evidence we *now* have; but there is a clear sense in which it may be false. Indeed, it may well lead to false predictions, and thus conflict with *OP*. But T_I, by hypothesis, does not lead to any false predictions. Still, the metaphysical realist claims – and it is just this claim that makes him a *metaphysical* as opposed to an empirical realist – that T_I may be, in reality, false. What is not knowable as true may nonetheless be true; what is epistemically most justifiable to believe may nonetheless be false, on this kind of realist view. The striking connection between issues and debates in the philosophy of science and issues and debates in the philosophy of mathematics is that this sort of realism runs into *precisely* the same difficulties that we saw Platonism run into. Let us pause to verify this.

Since the ideal theory T_I must, whatever other properties it may or

may not have, have the property of being *consistent*, it follows from the Gödel Completeness Theorem (whose proof, as all logicians know, is intimately related to one of Skolem's proofs of the Löwenheim-Skolem Theorem), that T_I has models. We shall assume that T_I contains a primitive or defined term denoting each member of S, the set of "observable things and events". The assumption that we made, that T_I agrees with *OP*, means that all those sentences about members of S which *OP* requires to be true are theorems of T_I. Thus if M is any model of T_I, M has to have a member corresponding to each member of S. We can even replace each member of M which corresponds to a member of S by that member of S itself, modifying the interpretation of the predicate letters accordingly, and obtain a model M' in which each term denoting a member of S in the "intended" interpretation does denote that member of S. Then the extension of each O-term in that model will be partially correct to the extent determined by *OP*: that is, everything that *OP* "says" is in the extension of $\underline{P}$ is in the extension of $\underline{P}$, and everything that *OP* "says" is in the extension of the complement of $\underline{P}$ is in the extension of the complement of $\underline{P}$, for each O-term, in any such model. In short, such a model is standard with respect to $P \restriction S$ (P restricted to S) for each O-term P.

Now, such a model satisfies all operational constraints, since it agrees with *OP*. It satisfies those theoretical constraints we would impose in the ideal limit of inquiry. So, once again, it looks as if any such model is "intended" – for what else could single out a model as "intended" than this? But if this is what it *is* to be an "intended model", T_I must be *true* – true in all intended models! The metaphysical realist's claim that even the ideal theory T_I might be false "in reality" seems to collapse into unintelligibility.

Of course, it might be contended that "true" does not follow from "true in all intended models". But "true" is the same as "true in the intended *interpretation*" (or "in *all* intended interpretations", if there may be more than one interpretation intended – or permitted – by the speaker), on any view. So to follow this line – which is, indeed, the right one, in my view – one needs to develop a theory on which interpretations are specified *other* than by specifying models.

Once again, an appeal to mysterious powers of the mind is made by some. Chisholm (following the tradition of Brentano) contends that the mind has a faculty of *referring to external objects* (or perhaps to external properties) which he calls by the good old name "intentionality". Once again most naturalistically minded philosophers (and, of course, psychologists), find the postulation of unexplained mental faculties unhelpful epistemology and almost certainly bad science as well.

There are two main tendencies in the philosophy of science (I hesitate to call them "views", because each tendency is represented by many different detailed views) about the way in which the reference of theoretical terms gets fixed. According to one tendency, which we may call the Ramsey tendency, and whose various versions constituted the received view for many years, theoretical terms come in batches or clumps. Each clump – for example, the clump consisting of the primitives of electromagnetic theory – is defined by a theory, in the sense that all the models of that theory which are standard on the observation terms count as intended models. The theory is "true" just in case it has such a model. (The "Ramsey sentence" of the theory is just the second-order sentence that asserts the existence of such a model.) A sophisticated version of this view, which amounts to relativizing the Ramsey sentence to an open set of "intended applications", has recently been advanced by Joseph Sneed.

The other tendency is the realist tendency. While realists differ among themselves even more than proponents of the (former) received view do, realists unite in agreeing that a theory may have a true Ramsey sentence and not be (in reality) true.

The first of the two tendencies I described, the Ramsey tendency, represented in the United States by the school of Rudolf Carnap, accepted the "relativity of theoretical notions", and abandoned the realist intuitions. The second tendency is more complex. Its, so to speak, conservative wing, represented by Chisholm, joins Plato and the ancients in postulating mysterious powers wherewith the mind "grasps" concepts, as we have already said. If we have more available with which to fix the intended model than merely theoretical and operational constraints, then the problem disappears. The radical pragmatist wing, represented, perhaps, by Quine, is willing to give up the intuition that T_I might be false "in reality". This radical wing is "realist" in the sense of being willing to assert that present-day science, taken more or less at face value (i.e., without philosophical reinterpretation) is at least approximately true; "realist" in the sense of regarding reference as trans-theoretic (a theory with a true Ramsey sentence may be false, because later inquiry may establish an incompatible theory as better); but not *metaphysical* realist. It is the moderate "center" of the realist tendency, the center that would like to hold on to metaphysical realism *without* postulating mysterious powers of the mind that is once again in deep trouble.

Pushing the problem back: the Skolemization of absolutely everything. We have seen that issues in the philosophy of science having to do with reference of theoretical terms and issues in the philosophy of mathe-

matics having to do with the problem of singling out a unique "intended model" for set theory are both connected with the Löwenheim-Skolem Theorem and its near relative, the Gödel Completeness Theorem. Issues having to do with reference also arise in philosophy in connection with sense data and material objects and, once again, these connect with the model-theoretic problems we have been discussing. (In some way, it really seems that the Skolem Paradox underlies the *characteristic* problems of 20th century philosophy.)

Although the philosopher John Austin and the psychologist Fred Skinner both tried to drive sense data out of existence, it seems to me that most philosophers and psychologists think that there are such things as *sensations,* or *qualia.* They may not be objects of perception, as was once thought (it is becoming increasingly fashionable to view them as states or conditions of the sentient subject, as Reichenbach long ago urged we should); we may not have incorrigible knowledge concerning them; they may be somewhat ill-defined entities rather than the perfectly sharp particulars they were once taken to be; but it seems reasonable to hold that they are part of the legitimate subject matter of cognitive psychology and philosophy and not mere pseudo-entities invented by bad psychology and bad philosophy.

Accepting this, and taking the operational constraint this time to be that we wish the ideal theory to correctly predict all sense data, it is easily seen that the previous argument can be repeated here, this time to show that (if the "intended" models are the ones which satisfy the operational and theoretical constraints we now have, or even the operational and theoretical constraints we would impose in some limit) then, either the present theory is "true", in the sense of being "true in all intended models", provided it leads to no false predictions about sense data, or else the ideal theory is "true". The first alternative corresponds to taking the theoretical constraints to be represented by current theory; the second alternative corresponds to taking the theoretical constraints to be represented by the ideal theory. This time, however, it will be the case that even terms referring to ordinary material objects – terms like 'cat' and 'dog' – get differently interpreted in the different "intended" models. It seems, this time, as if we cannot even refer to ordinary middle sized physical objects except as formal constructs variously interpreted in various models.

Moreover, if we agree with Wittgenstein that the *similarity relation* between sense data we have at different times is not itself something present to my mind – that "fixing one's attention" on a sense datum and thinking "by 'red' I mean whatever is like *this*" does not really pick out any relation of similarity at all – and make the natural move of supposing that the intended models of my language when I now and in the future talk of the sense data I had at some past time t_0 are singled out by operational and theoretical constraints, then, again, it will turn out that my *past* sense data are mere formal constructs which get differently interpreted in various models. If we further agree with Wittgenstein that the notion of truth requires a *public* language (or requires at least states of the self at more than one time – that a "private language for one specious present" makes no sense), then even my *present* sense data are in this same boat.... In short, one can "Skolemize" absolutely everything. It seems to be absolutely impossible to fix a determinate reference (without appeal to nonnatural mental powers) for *any* term at all. If we apply the argument to the very metalanguage we use to talk about the predicament...?

The same problem has even surfaced recently in the field of cognitive psychology. The standard model for the brain/mind in this field is the modern computing machine. This computing machine is thought of as having something analogous to a formalized language in which it computes. (This hypothetical brain language has even received a name – "mentalese".) What makes the model of cognitive psychology a *cognitive* model is that "mentalese" is thought to be a medium whereby the brain constructs an *internal representation* of the external world. This idea runs immediately into the following problem: if "mentalese" is to be a vehicle for describing the external world, then the various predicate letters must have extensions which are sets of external things (or sets of *n*tuples of external things). But if the way "mentalese" is "understood" by the deep structures in the brain that compute, record, etc. in this "language" is *via* what artificial intelligence people call "procedural semantics" – that is, if the brain's *program for using* "mentalese" comprises its entire "understanding" of "mentalese" – where the program for using "mentalese", like any program, refers only to what is *inside* the computer – then how do *extensions* ever come into the picture at all? In the terminology I have been employing in this address, the problem is this: if the extension of predicates in "mentalese" is fixed by the theoretical and operational constraints "hard wired in" to the brain, or even by theoretical and operational constraints that it evolves in the course of inquiry, then these will not fix a *determinate* extension for any predicate. If thinking is ultimately done in "mentalese", then *no concept we have will have a determinate extension.* Or so it seems.

The bearing of causal theories of reference. The term "causal theory of reference" was originally applied to my theory of the reference of natural kind terms and Kripke's theory of the reference of proper names. These

theories did not attempt to *define* reference, but rather attempted to say something about how reference is fixed, if it is not fixed by associating definite descriptions with the terms and names in question. Kripke and I argued that the intention to preserve reference through a historical chain of uses and the intention to cooperate socially in the fixing of reference make it possible to use terms successfully to refer although no one definite description is associated with any term by all speakers who use that term. These theories assume that individuals can be singled out for the purpose of a "naming ceremony" and that inferences to the existence of definite theoretical entities (to which names can then be attached) can be successfully made. Thus these theories did not address the question as to how any term can acquire a determinate reference (or any gesture, e.g., pointing – of course, the "reference" of gestures is just as problematic as the reference of terms, if not more so). Recently, however, it has been suggested by various authors that some account can be given of how at least some basic sorts of terms refer in terms of the notion of a "causal chain". In one version (cf. Evans 1973: 187–208), a version strikingly reminiscent of the theories of Ockham and other 14th century logicians, it is held that a term refers to "the dominant source" of the beliefs that contain the term. Assuming we can circumvent the problem that the dominant cause of our beliefs concerning *electrons* may well be *textbooks*,[4] it is important to notice that even if a *correct* view of this kind can be elaborated, it will do nothing to resolve the problem we have been discussing.

The problem is that adding to our hypothetical formalized language of science a body of theory titled "causal theory of reference" *is* just adding more *theory*. But Skolem's argument, and our extensions of it, are not affected by enlarging the theory. Indeed, you can even take the theory to consist of *all true sentences*, and there will be many models – models differing on the extension of every term not fixed by *OP* (or whatever you take *OP* to be in a given context) – which satisfy the entire theory. If "refers" can be defined in terms of some causal predicate or predicates in the metalanguage of our theory, then, since each model of the object language extends in an obvious way to a corresponding model of the metalanguage, it will turn out that, *in each model M, reference$_M$* is definable in terms of *causes$_M$*; but, unless the word 'causes' (or whatever the causal predicate or predicates may be) is already glued to one definite relation with metaphysical glue, this does not fix a determinate extension for 'refers' at all.

[4] Evans handles this case by saying that there are appropriateness conditions on the type of causal chain which must exist between the item referred to and the speaker's body of information.

This is not to say that the construction of such a theory would be worthless as philosophy or as natural science. The program of cognitive psychology already alluded to – the program of describing our brains as computers which construct an "internal representation of the environment" seems to require that "mentalese" utterances be, in some cases at least, describable as the causal product of devices in the brain and nervous system which "transduce" information from the environment, and such a description might well be what the causal theorists are looking for. The program of realism in the philosophy of science – of *empirical* realism, not metaphysical realism – is to show that scientific theories can be regarded as better and better representations of an objective world with which we are interacting; if such a view is to be part of science itself, as empirical realists contend it should be, then the interactions with the world by means of which this representation is formed and modified must themselves be part of the subject matter of the representation. But the problem as to how the *whole representation*, including the empirical theory of knowledge that is a part of it, can determinately refer is not a problem that can be solved by developing more and better empirical theory.

Ideal theories and truth. One reaction to the problem I have posed would be to say: there are many ideal theories in the sense of theories which satisfy the operational constraints, and in addition have all the virtues (simplicity, coherence, containing the axiom of choice, whatever) that humans like to demand. But there are no "facts of the matter" not reflected in constraints on ideal theories in this sense. Therefore, what is really true is what is common to all such ideal theories; what is really false is what they all deny; all other statements are neither true nor false.

Such a reaction would lead to too few truths, however. It may well be that there are rational beings – even rational human species – which do not employ our color predicates, or who do not employ the predicate "person", or who do not employ the predicate "earthquake" (cf. Wiggins 1977). I see no reason to conclude from this that *our* talk of red things, or of persons, or of earthquakes, lacks truth value. If there are many ideal theories (and if "ideal" is itself a somewhat interest-relative notion), if there are many theories which (given appropriate circumstances) it is perfectly rational to accept, then it seems better to say that, insofar as these theories say different (and sometimes, apparently incompatible) things, that some facts are "soft" in the sense of depending for their truth value on the speaker, the circumstances of utterance, etc. This is what we have to say in any case about cases of ordinary vagueness, about ordinary causal talk, etc. It is what we say about apparently

incompatible statements of simultaneity in the special theory of relativity. To grant that there is more than one true version of reality is not to deny that some versions are false.

It may be, of course, that there *are* some truths that *any* species of rational inquirers would eventually acknowledge. (On the other hand, the set of these may be empty, or almost empty.) But to say that *by definition* these are all the truths there are is to redefine the notion in a highly restrictive way. (It also assumes that the notion of an "ideal theory" is perfectly clear; an assumption which seems plainly false.)

Intuitionism. It is a striking fact that this entire problem does *not* arise for the standpoint of mathematical intuitionism. This would not be a surprise to Skolem: it was precisely his conclusion that "most mathematicians want mathematics to deal, ultimately, with performable computing operations and not to consist of formal propositions about objects called this or that."

In intuitionism, knowing the meaning of a sentence or predicate consists in associating the sentence or predicate with a procedure which enables one to recognize when one has a proof that the sentence is constructively true (i.e., that it is possible to carry out the constructions that the sentence asserts can be carried out) or that the predicate applies to a certain entity (i.e., that a certain full sentence of the predicate is constructively true). The most striking thing about this standpoint is that the *classical notion of truth is nowhere used* – the semantics is entirely given in terms of the notion of "constructive proof", *including the semantics of "constructive proof" itself.*

Of course, the intuitionists do not think that "constructive proof" can be formalized, or that "mental constructions" can be identified with operations in our *brains*. Generally, they assume a strongly intentionalist and *a prioristic* posture in philosophy – that is, they assume the existence of mental entities called "meanings" and of a special faculty of intuiting constructive relations between these entities. These are not the aspects of intuitionism I shall be concerned with. Rather I wish to look on intuitionism as an example of what Michael Dummett has called "non-realist semantics" – that is, a semantic theory which holds that *a language is completely understood when a verification procedure is suitably mastered,* and not when truth conditions (in the classical sense) are learned.

The problem with realist semantics – truth-conditional semantics – as Dummett has emphasized, is that if we hold that the understanding of the sentences of, say, set theory consists in our knowledge of their "truth conditions", then how can we possibly say what *that* knowledge in turn consists in? (It cannot, as we have just seen, consist in the use of language or "mentalese" under the control of operational plus theoretical constraints, be they fixed or evolving, since such constraints are too weak to provide a determinate extension for the terms, and it is this that the realist wants.)

If, however, the understanding of the sentences of a mathematical theory consists in the mastery of verification procedures (which need not be fixed once and for all – we can allow a certain amount of "creativity"), then a mathematical theory can be completely understood, and this understanding does not presuppose the notion of a "model" at all, let alone an "intended model".

Nor does the intuitionist (or, more generally, the "nonrealist" semanticist) have to foreswear *forever* the notion of a model. He has to foreswear reference to models in his account of *understanding*; but, once he has succeeded in understanding a rich enough language to serve as a metalanguage for some theory T (which may itself be simply a sublanguage of the metalanguage, in the familiar way), he can define 'true in T' à la Tarski, he can talk about "models" for T, etc. He can even define 'reference' or ('satisfaction') exactly as Tarski did.

Does the whole "Skolem Paradox" arise again to plague him at this stage? The answer is that it does not. To see why it does not, one has to realize what the "existence of a model" means in *constructive* mathematics.

"Objects" in constructive mathematics are *given through descriptions.* Those descriptions do not have to be mysteriously attached to those objects by some nonnatural process (or by metaphysical glue). Rather the possibility of *proving* that a certain construction (the "sense", so to speak, of the description of the model) has certain constructive properties is what is asserted and *all* that is asserted by saying the model "exists". In short, ***reference is given through sense, and sense is given through verification-procedures and not through truth-conditions.*** The "gap" between our theory and the "objects" simply disappears – or, rather, it never appears in the first place.

Intuitionism liberalized. It is not my aim, however, to try to convert my audience to intuitionism. Set theory may not be the "paradise" Cantor thought it was, but it is not such a bad neighborhood that I want to leave of my own accord, either. Can we separate the philosophical idea behind intuitionism, the idea of "nonrealist" semantics, from the restrictions and prohibitions that the historic intuitionists wished to impose upon mathematics?

The answer is that we can. First, as to set theory: the objection to *impredicativity,* which is the intuitionist ground for rejecting much of

classical set theory, has little or no connection with the insistence upon verificationism itself. Indeed, intuitionist mathematics is itself "impredicative", inasmuch as the intuitionist notion of constructive proof presupposes constructive proofs which refer to the totality of *all* constructive proofs.

Second, as to the propositional calculus: it is well known that the classical connectives can be reintroduced into an intuitionist theory by reinterpretation. The important thing is not whether one uses "classical propositional calculus" or not, but how one *understands* the logic if one does use it. Using classical logic as an intuitionist would understand it, means, for example, keeping track of when a disjunction is selective (i.e., one of the disjuncts is constructively provable), and when it is nonselective; but this does not seem like too bad an idea.

In short, while intuitionism may go with a greater interest in constructive mathematics, a liberalized version of the intuitionist standpoint need not rule out "classical" mathematics as either illegitimate or unintelligible. What about the language of empirical science? Here there are greater difficulties. Intuitionist logic is given in terms of a notion of *proof,* and proof is supposed to be a *permanent* feature of statements. Moreover, proof is nonholistic; there is such a thing as the proof (in either the classical or the constructive sense) of an isolated mathematical statement. But verification in empirical science is a matter of degree, not a "yes-or-no" affair; even if we made it a "yes-or-no" affair in some arbitrary way, verification is a property of empirical sentences that can be *lost;* in general the "unit of verification" in empirical science is the theory and not the isolated statement.

These difficulties show that sticking to the intuitionist standpoint, however liberalized, would be a bad idea in the context of formalizing empirical science. But they are not incompatible with "nonrealist" semantics. The crucial question is this: do we think of the *understanding* of the language as consisting in the fact that speakers possess (collectively if not individually) an evolving network of verification procedures, or as consisting in their possession of a set of "truth conditions"? If we choose the first alternative, the alternative of "nonrealist" semantics, then the "gap" between words and world, between our *use* of the language and its "objects", never appears.[5] Moreover, the "nonrealist"

[5]To the suggestion that we identify truth with being verified, or accepted, or accepted in the long run, it may be objected that a person could reasonably, and possibly truly, make the assertion:

> A; but it could have been the case that A and our scientific development differ in such a way to make $\bar{A}$ part of the ideal theory accepted in the long run; in that circumstance, it would have been the case that A but it was not true that A.

semantics is not *inconsistent* with realist semantics; it is simply *prior* to it, in the sense that it is the "nonrealist" semantics that must be internalized if the language is to be understood.

Even if it is not inconsistent with realist semantics, taking the nonrealist semantics as our picture of how the language is understood undoubtedly will affect the way we view questions about reality and truth. For one thing, verification in empirical science (and, to a lesser extent, in mathematics as well, perhaps) sometimes depends on what we before called "decision" or "convention". Thus facts may, on this picture, depend on our interests, saliencies and decisions. There will be many "soft facts". (Perhaps whether $V=L$ or not is a "soft fact".) I cannot, myself, regret this. If appearance and reality end up being endpoints on a continuum rather than being the two halves of a monster Dedekind cut in all we conceive and do not conceive, it seems to me that philosophy will be much better off. The search for the "furniture of the Universe" will have ended with the discovery that the Universe is not a furnished room.

Where did we go wrong? – The problem solved. What Skolem really pointed out is this: no interesting theory (in the sense of first-order theory) can, in and of itself, determine its own objects up to isomorphism. Skolem's argument can be extended as we saw, to show that if theoretical constraints do not determine reference, then the addition of operational constraints will not do it either. It is at this point that reference itself begins to seem "occult"; that it begins to seem that one cannot be any kind of a realist without being a believer in nonnatural mental powers. Many moves have been made in response to this predicament, as we noted above. Some have proposed that *second-order* formalizations are the solution, at least for mathematics; but the "intended" interpretation of the second-order formalism is not fixed by the use of the formalism (the formalism itself admits so-called "Henkin models", i.e., models in which the second-order variables fail to range over the *full* power set of the universe of individuals), and it becomes necessary to

This argument is fallacious, however, because the different "scientific development" means here the choice of a different version; we cannot assume the *sentence* ⌜A⌝ has a fixed meaning independent of what version we accept.

In fact the same problem can confront a metaphysical realist. Realists also have to recognize that there are cases in which the reference of a term depends on which theory one accepts, so that A can be a true sentence if T_1 is accepted and a false one if T_2 is accepted, where T_1 and T_2 are both true theories. But then imagine someone saying,

> A; but it could have been the case that our A and our scientific development differ in such a way that T_2 was accepted. In that case, it would have been the case that A but A would not have been true.

attribute to the mind special powers of "grasping second-order notions". Some have proposed to accept the conclusion that mathematical language is only partially interpreted, and likewise for the language we use to speak of "theoretical entities" in empirical science; but then are "ordinary material objects" any better off? Are sense data better off? Both Platonism and phenomenalism have run rampant at different times and in different places in response to this predicament.

The problem, however, lies with the predicament itself. The predicament only *is* a predicament because we did two things: first, we gave an account of understanding the language in terms of programs and procedures for *using* the language (what else?); then, secondly, we asked what the possible "models" for the language were, thinking of the models as existing "out there" *independent of any description*. At this point, something really weird had already happened, had we stopped to notice. On any view, the understanding of the language must determine the reference of the terms, or, rather, must determine the reference given the context of use. If the use, even in a fixed context, does not determine reference, then use is not understanding. The language, on the perspective we talked ourselves into, has a full program of use; but it still lacks an *interpretation*.

This is the fatal step. To adopt a theory of meaning according to which a language whose whole use is specified still lacks something – viz. its "interpretation" – is to accept a problem which *can* only have crazy solutions. To speak as if *this* were my problem, "I know how to use my language, but, now, how shall I single out an interpretation?" is to speak nonsense. Either the use *already* fixes the "interpretation" or *nothing* can.

Nor do "causal theories of reference", etc., help. Basically, trying to get out of this predicament by *these* means is hoping that the *world* will pick one definite extension for each of our terms even if *we* cannot. But the world does not pick models or interpret languages. *We* interpret our languages or nothing does.

We need, therefore, a standpoint which links use and reference in just the way that the metaphysical realist standpoint refuses to do. The standpoint of "non-realist semantics" is precisely that standpoint. From that standpoint, it is trivial to say that a model in which, as it might be, the set of cats and the set of dogs are permuted (i.e., 'cat' is assigned the set of dogs as its extension, and 'dog' is assigned the set of cats) is "unintended" even if corresponding adjustments in the extensions of all the other predicates make it end up that the operational and theoretical constraints of total science or total belief are all "preserved". Such a model would be unintended *because we do not intend the word 'cat' to refer to* *dogs*. From the metaphysical realist standpoint, this answer does not work; it just pushes the question back to the metalanguage. The axiom of the metalanguage, "'cat' refers to cats" cannot rule out such an unintended interpretation of the object language, unless the metalanguage itself already has had *its* intended interpretation singled out; but we are in the same predicament with respect to the metalanguage that we are in with respect to the object language, from that standpoint, so all is in vain. However, from the viewpoint of "nonrealist" semantics, the metalanguage is completely understood, and so is the object language. So we can *say and understand*, "'cat' refers to cats". Even though the model referred to satisfies the theory, etc., it is "unintended"; we recognize that it is unintended *from the description through which it is given* (as in the intuitionist case). Models are not lost noumenal waifs looking for someone to name them; they are constructions within our theory itself, and they have names from birth.

PART IV

The concept of set

Russell's mathematical logic

KURT GÖDEL

Mathematical logic, which is nothing else but a precise and complete formulation of formal logic, has two quite different aspects. On the one hand, it is a section of Mathematics treating of classes, relations, combinations of symbols, etc., instead of numbers, functions, geometric figures, etc. On the other hand, it is a science prior to all others, which contains the ideas and principles underlying all sciences. It was in this second sense that Mathematical Logic was first conceived by Leibniz in his *Characteristica universalis,* of which it would have formed a central part. But it was almost two centuries after his death before his idea of a logical calculus really sufficient for the kind of reasoning occurring in the exact sciences was put into effect (in some form at least, if not the one Leibniz had in mind) by Frege and Peano.[1] Frege was chiefly interested in the analysis of thought and used his calculus in the first place for deriving arithmetic from pure logic. Peano, on the other hand, was more interested in its applications within mathematics and created an elegant and flexible symbolism, which permits expressing even the most complicated mathematical theorems in a perfectly precise and often very concise manner by single formulas.

It was in this line of thought of Frege and Peano that Russell's work set in. Frege, in consequence of his painstaking analysis of the proofs, had not gotten beyond the most elementary properties of the series of integers, while Peano had accomplished a big collection of mathematical theorems expressed in the new symbolism, but without proofs. It was

Reprinted with the kind permission of the author, editor, and publisher from Paul A. Schilpp, ed., *The Philosophy of Bertrand Russell,* The Library of Living Philosophers, Evanston, Ill. (Evanston & Chicago: Northwestern University, 1944), pp. 125–53. The author asked to note (1) that since the original publication of this paper advances have been made in some of the problems discussed and that the formulations given could be improved in several places, and (2) that the term "constructivistic" in this paper is used for a strictly anti-realistic kind of constructivism. Its meaning, therefore, is not identical with that used in current discussions on the foundations of mathematics. If applied to the actual development of logic and mathematics it is equivalent with a certain kind of "predicativity" and hence different both from "intuitionistically admissible" and from "constructive" in the sense of the Hilbert School.

[1] Frege has doubtless the priority, since his first publication about the subject, which already contains all the essentials, appeared ten years before Peano's.

only in *Principia Mathematica* that full use was made of the new method for actually deriving large parts of mathematics from a very few logical concepts and axioms. In addition, the young science was enriched by a new instrument, the abstract theory of relations. The calculus of relations had been developed before by Peirce and Schröder, but only with certain restrictions and in too close analogy with the algebra of numbers. In *Principia* not only Cantor's set theory but also ordinary arithmetic and the theory of measurement are treated from this abstract relational standpoint.

It is to be regretted that this first comprehensive and thorough going presentation of a mathematical logic and the derivation of Mathematics from it is so greatly lacking in formal precision in the foundations (contained in *1–*21 of *Principia*), that it presents in this respect a considerable step backwards as compared with Frege. What is missing, above all, is a precise statement of the syntax of the formalism. Syntactical considerations are omitted even in cases where they are necessary for the cogency of the proofs, in particular in connection with the "incomplete symbols." These are introduced not by explicit definitions, but by rules describing how sentences containing them are to be translated into sentences not containing them. In order to be sure, however, that (or for what expressions) this translation is possible and uniquely determined and that (or to what extent) the rules of inference apply also to the new kind of expressions, it is necessary to have a survey of all possible expressions, and this can be furnished only by syntactical considerations. The matter is especially doubtful for the rule of substitution and of replacing defined symbols by their *definiens*. If this latter rule is applied to expressions containing other defined symbols it requires that the order of elimination of these be indifferent. This however is by no means always the case ($\varphi!\hat{u} = \hat{u}[\varphi!u]$, e.g., is a counter-example). In *Principia* such eliminations are always carried out by substitutions in the theorems corresponding to the definitions, so that it is chiefly the rule of substitution which would have to be proved.

I do not want, however, to go into any more details about either the formalism or the mathematical content of *Principia*,[2] but want to devote the subsequent portion of this essay to Russell's work concerning the analysis of the concepts and axioms underlying Mathematical Logic. In this field Russell has produced a great number of interesting ideas some of which are presented most clearly (or are contained only) in his earlier writings. I shall therefore frequently refer also to these earlier writings,

[2]Cf. in this respect Quine 1941.

although their content may partly disagree with Russell's present standpoint.

What strikes one as surprising in this field is Russell's pronouncedly realistic attitude, which manifests itself in many passages of his writings. "Logic is concerned with the real world just as truly as zoology, though with its more abstract and general features," he says, e.g., in his *Introduction to Mathematical Philosophy* (edition of 1920, p. 169). It is true, however, that this attitude has been gradually decreasing in the course of time[3] and also that it always was stronger in theory than in practice. When he started on a concrete problem, the objects to be analyzed, (e.g., the classes or propositions) soon for the most part turned into "logical fictions." Though perhaps this need not necessarily mean [according to the sense in which Russell uses this term] that these things do not exist, but only that we have no direct perception of them.

The analogy between mathematics and a natural science is enlarged upon by Russell also in another respect (in one of his earlier writings). He compares the axioms of logic and mathematics with the laws of nature and logical evidence with sense perception, so that the axioms need not necessarily be evident in themselves, but rather their justification lies (exactly as in physics) in the fact that they make it possible for these "sense perceptions" to be deduced; which of course would not exclude that they also have a kind of intrinsic plausibility similar to that in physics. I think that (provided "evidence" is understood in a sufficiently strict sense) this view has been largely justified by subsequent developments, and it is to be expected that it will be still more so in the future. It has turned out that (under the assumption that modern mathematics is consistent) the solution of certain arithmetical problems requires the use of assumptions essentially transcending arithmetic, i.e., the domain of the kind of elementary indisputable evidence that may be most fittingly compared with sense perception. Furthermore it seems likely that for deciding certain questions of abstract set theory and even for certain related questions of the theory of real numbers new axioms based on some hitherto unknown idea will be necessary. Perhaps also the apparently unsurmountable difficulties which some other mathematical problems have been presenting for many years are due to the fact that the necessary axioms have not yet been found. Of course, under these circumstances mathematics may lose a good deal of its "absolute certainty;" but, under the influence of the modern criticism of the foundations, this has already happened to a large extent. There is some resemblance

[3]The above quoted passage was left out in the later editions of the *Introduction*.

between this conception of Russell and Hilbert's "supplementing the data of mathematical intuition" by such axioms as, e.g., the law of excluded middle which are not given by intuition according to Hilbert's view; the borderline however between data and assumptions would seem to lie in different places according to whether we follow Hilbert or Russell.

An interesting example of Russell's analysis of the fundamental logical concepts is his treatment of the definite article "the". The problem is: what do the so-called descriptive phrases (i.e., phrases as, e.g., "the author of *Waverley*" or "the king of England") denote or signify[4] and what is the meaning of sentences in which they occur? The apparently obvious answer that, e.g., "the author of *Waverley*" signifies Walter Scott, leads to unexpected difficulties. For, if we admit the further apparently obvious axiom, that the signification of a composite expression, containing constituents which have themselves a signification, depends only on the signification of these constituents (not on the manner in which this signification is expressed), then it follows that the sentence "Scott is the author of *Waverley*" signifies the same thing as "Scott is Scott"; and this again leads almost inevitably to the conclusion that all true sentences have the same signification (as well as all false ones).[5] Frege actually drew this conclusion; and he meant it in an almost metaphysical sense, reminding one somewhat of the Eleatic doctrine of the "One." "The True" – according to Frege's view – is analyzed by us in different ways in different propositions; "the True" being the name he uses for the common signification of all true propositions (cf. 1892b: 35).

Now according to Russell, what corresponds to sentences in the outer world is facts. However, he avoids the term "signify" or "denote" and uses "indicate" instead (in his earlier papers he uses "express" or "being a symbol for"), because he holds that the relation between a sentence and a fact is quite different from that of a name to the thing named. Furthermore, he uses "denote" (instead of "signify") for the relation between things and names, so that "denote" and "indicate" together would correspond to Frege's "*bedeuten*". So, according to Russell's

[4]I use the term "signify" in the sequel because it corresponds to the German word "*bedeuten*" which Frege, who first treated the question under consideration, used in this connection.

[5]The only further assumptions one would need in order to obtain a rigorous proof would be: (1) that "$\varphi(a)$" and the proposition "a is the object which has the property φ and is identical with a" mean the same thing and (2) that every proposition "speaks about something," i.e., can be brought to the form $\varphi(a)$. Furthermore one would have to use the fact that for any two objects a, b, there exists a true proposition of the form $\varphi(a,b)$ as, e.g., $a \neq b$ or $a = a \cdot b = b$.

terminology and view, true sentences "indicate" facts and, correspondingly, false ones indicate nothing.[6] Hence Frege's theory would in a sense apply to false sentences, since they all indicate the same thing, namely nothing. But different true sentences may indicate many different things. Therefore this view concerning sentences makes it necessary either to drop the above-mentioned principle about the signification (i.e., in Russell's terminology the corresponding one about the denotation and indication) of composite expressions or to deny that a descriptive phrase denotes the object described. Russell did the latter[7] by taking the viewpoint that a descriptive phrase denotes nothing at all but has meaning only in context; for example, the sentence "the author of *Waverley* is Scotch", is defined to mean: "There exists exactly one entity who wrote *Waverley* and whoever wrote *Waverley* is Scotch." This means that a sentence involving the phrase "the author of *Waverley*" does not (strictly speaking) assert anything about Scott (since it contains no constituent denoting Scott), but is only a roundabout way of asserting something about the concepts occurring in the descriptive phrase. Russell adduces chiefly two arguments in favor of this view, namely (1) that a descriptive phrase may be meaningfully employed even if the object described does not exist (e.g., in the sentence: "The present king of France does not exist"). (2) That one may very well understand a sentence containing a descriptive phrase without being acquainted with the object described; whereas it seems impossible to understand a sentence without being acquainted with the objects about which something is being asserted. The fact that Russell does not consider this whole question of the interpretation of descriptions as a matter of mere linguistic conventions, but rather as a question of right and wrong, is another example of his realistic attitude, unless perhaps he was aiming at a merely psychological investigation of the actual processes of thought. As to the question in the logical sense, I cannot help feeling that the problem raised by Frege's puzzling conclusion has only been evaded by Russell's theory of descriptions and that there is something behind it which is not yet completely understood.

There seems to be one purely formal respect in which one may give preference to Russell's theory of descriptions. By defining the meaning

[6]From the indication (*Bedeutung*) of a sentence is to be distinguished what Frege called its meaning (*Sinn*) which is the conceptual correlate of the objectively existing fact (or "the True"). This one should expect to be in Russell's theory a possible fact (or rather the possibility of a fact), which would exist also in the case of a false proposition. But Russell, as he says, could never believe that such "curious shadowy" things really exist. Thirdly, there is also the psychological correlate of the fact which is called "signification" and understood to be the corresponding belief in Russell's latest book. "Sentence" in contradistinction to "proposition" is used to denote the mere combination of symbols.

[7]He made no explicit statement about the former; but it seems it would hold for the logical system of *Principia*, though perhaps more or less vacuously.

of sentences involving descriptions in the above manner, he avoids in his logical system any axioms about the particle "the", i.e., the analyticity of the theorems about "the" is made explicit; they can be shown to follow from the explicit definition of the meaning of sentences involving "the". Frege, on the contrary, has to assume an axiom about "the", which of course is also analytic, but only in the implicit sense that it follows from the meaning of the undefined terms. Closer examination, however, shows that this advantage of Russell's theory over Frege's subsists only as long as one interprets definitions as mere typographical abbreviations, not as introducing names for objects described by the definitions, a feature which is common to Frege and Russell.

I pass now to the most important of Russell's investigations in the field of the analysis of the concepts of formal logic, namely those concerning the logical paradoxes and their solution. By analyzing the paradoxes to which Cantor's set theory had led, he freed them from all mathematical technicalities, thus bringing to light the amazing fact that our logical intuitions (i.e., intuitions concerning such notions as: truth, concept, being, class, etc.) are self-contradictory. He then investigated where and how these common-sense assumptions of logic are to be corrected and came to the conclusion that the erroneous axiom consists in assuming that for every propositional function there exists the class of objects satisfying it, or that every propositional function exists "as a separate entity;"[8] by which is meant something separable from the argument (the idea being that propositional functions are abstracted from propositions which are primarily given) and also something distinct from the combination of symbols expressing the propositional function; it is then what one may call the notion or concept defined by it.[9] The existence of this concept already suffices for the paradoxes in their "intensional" form, where the concept of "not applying to itself" takes the place of Russell's paradoxical class.

Rejecting the existence of a class or concept in general, it remains to determine under what further hypotheses (concerning the propositional function) these entities do exist. Russell pointed out (1907: 29) two possible directions in which one may look for such a criterion, which he

[8]In Russell 1907: 29. If one wants to bring such paradoxes as "the liar" under this viewpoint, universal (and existential) propositions must be considered to involve the class of objects to which they refer.

[9]"Propositional function" (without the clause "as a separate entity") may be understood to mean a proposition in which one or several constituents are designated as arguments. One might think that the pair consisting of the proposition and the argument could then for all purposes play the rôle of the "propositional function as a separate entity," but it is to be noted that this pair (as one entity) is again a set or a concept and therefore need not exist.

called the zig-zag theory and the theory of limitation of size, respectively, and which might perhaps more significantly be called the intensional and the extensional theory. The second one would make the existence of a class or concept depend on the extension of the propositional function (requiring that it be not too big), the first one on its content or meaning (requiring a certain kind of "simplicity," the precise formulation of which would be the problem).

The most characteristic feature of the second (as opposed to the first) would consist in the non-existence of the universal class or (in the intensional interpretation) of the notion of "something" in an unrestricted sense. Axiomatic set theory as later developed by Zermelo and others can be considered as an elaboration of this idea as far as classes are concerned.[10] In particular the phrase "not too big" can be specified (as was shown by J. v. Neumann 1929: 227) to mean: not equivalent with the universe of all things, or, to be more exact, a propositional function can be assumed to determine a class when and only when there exists no relation (in intension, i.e., a propositional function with two variables) which associates in a one-to-one manner with each object, an object satisfying the propositional function and vice versa. This criterion, however, does not appear as the basis of the theory but as a consequence of the axioms and inversely can replace two of the axioms (the axiom of replacement and that of choice).

For the second of Russell's suggestions too, i.e., for the zig-zag theory, there has recently been set up a logical system which shares some essential features with this scheme, namely Quine's system (cf. 1937: 70). It is, moreover, not unlikely that there are other interesting possibilities along these lines.

Russell's own subsequent work concerning the solution of the paradoxes did not go in either of the two afore-mentioned directions pointed out by himself, but was largely based on a more radical idea, the "no-class theory," according to which classes or concepts *never* exist as real objects, and sentences containing these terms are meaningful only to such an extent as they can be interpreted as a *façon de parler*, a manner of speaking about other things (cf. p. [460]). Since in *Principia* and elsewhere, however, he formulated certain principles discovered in the course of the development of this theory as general logical principles without mentioning any longer their dependence on the no-class theory, I am going to treat of these principles first.

I mean in particular the vicious circle principle, which forbids a certain

[10]The intensional paradoxes can be dealt with, e.g., by the theory of simple types or the ramified hierarchy, which do not involve any undesirable restrictions if applied to concepts only and not to sets.

kind of "circularity" which is made responsible for the paradoxes. The fallacy in these, so it is contended, consists in the circumstance that one defines (or tacitly assumes) totalities, whose existence would entail the existence of certain new elements of the same totality, namely elements definable only in terms of the whole totality. This led to the formulation of a principle which says that no totality can contain members definable only in terms of this totality, or members involving or presupposing this totality [vicious circle principle]. In order to make this principle applicable to the intensional paradoxes, still another principle had to be assumed, namely that "every propositional function presupposes the totality of its values" and therefore evidently also the totality of its possible arguments (cf. Whitehead and Russell 1910–13, 2: 39). [Otherwise the concept of "not applying to itself" would presuppose no totality (since it involves no quantifications),[11] and the vicious circle principle would not prevent its application to itself.] A corresponding vicious circle principle for propositional functions which says that nothing defined in terms of a propositional function can be a possible argument of this function is then a consequence (cf. Whitehead and Russell 1910–13, 1: 47, section 4). The logical system to which one is led on the basis of these principles is the theory of orders in the form adopted, e.g., in the first edition of *Principia*, according to which a propositional function which either contains quantifications referring to propositional functions of order n or can be meaningfully asserted of propositional functions of order n is at least of order $n+1$, and the range of significance of a propositional function as well as the range of a quantifier must always be confined to a definite order.

In the second edition of *Principia*, however, it is stated in the Introduction (pp. XI and XII) that "in a limited sense" also functions of a higher order than the predicate itself (therefore also functions defined in terms of the predicate as, e.g., in $p'\kappa \in \kappa$) can appear as arguments of a predicate of functions; and in appendix B such things occur constantly. This means that the vicious circle principle for propositional functions is virtually dropped. This change is connected with the new axiom that functions can occur in propositions only "through their values," i.e., extensionally, which has the consequence that any propositional function can take as an argument any function of appropriate type, whose extension is defined (no matter what order of quantifiers is used in the definition of this extension). There is no doubt that these things are quite unobjection-

[11]Quantifiers are the two symbols $(\exists x)$ and (x) meaning respectively, "there exists an object x" and "for all objects x." The totality of objects x to which they refer is called their range.

able even from the constructive standpoint (see below and p. [456]), provided that quantifiers are always restricted to definite orders. The paradoxes are avoided by the theory of simple types,[12] which in *Principia* is combined with the theory of orders (giving as a result the "ramified hierarchy") but is entirely independent of it and has nothing to do with the vicious circle principle (cf. pp. [464–5]).

Now as to the vicious circle principle proper, as formulated on p. [454], it is first to be remarked that, corresponding to the phrases "definable only in terms of," "involving," and "presupposing," we have really three different principles, the second and third being much more plausible than the first. It is the first form which is of particular interest, because only this one makes impredicative definitions[13] impossible and thereby destroys the derivation of mathematics from logic, effected by Dedekind and Frege, and a good deal of modern mathematics itself. It is demonstrable that the formalism of classical mathematics does not satisfy the vicious circle principle in its first form, since the axioms imply the existence of real numbers definable in this formalism only by reference to all real numbers. Since classical mathematics can be built up on the basis of *Principia* (including the axiom of reducibility), it follows that even *Principia* (in the first edition) does not satisfy the vicious circle principle in the first form, if "definable" means "definable within the system" and no methods of defining outside the system (or outside other systems of classical mathematics) are known except such as involve still more comprehensive totalities than those occurring in the systems.

I would consider this rather as a proof that the vicious circle principle is false than that classical mathematics is false, and this is indeed plausible also on its own account. For, first of all one may, on good grounds, deny that reference to a totality necessarily implies reference to all single elements of it or, in other words, that "all" means the same as an infinite logical conjunction. One may, e.g., follow Langford's (1927: 599) and Carnap's (1931: 103 [51 in this volume], and 1937: 162) suggestion to

[12]By the theory of simple types I mean the doctrine which says that the objects of thought (or, in another interpretation, the symbolic expressions) are divided into types, namely: individuals, properties of individuals, relations between individuals, properties of such relations, etc. (with a similar hierarchy for extensions), and that sentences of the form: "a has the property φ," "b bears the Relation R to c," etc. are meaningless, if a, b, c, R, φ are not of types fitting together. Mixed types (such as the class of all classes of finite types) are excluded. That the theory of simple types suffices for avoiding also the epistemological paradoxes is shown by a closer analysis of these. (Cf. Ramsey 1926a and Tarski 1935b: 399.)

[13]These are definitions of an object α by reference to a totality to which α itself (and perhaps also things definable only in terms of α) belong. As, e.g., if one defines a class α as the intersection of all classes satisfying a certain condition φ and then concludes that α is a subset also of such classes u as are defined in terms of α (provided they satisfy φ).

interpret "all" as meaning analyticity or necessity or demonstrability. There are difficulties in this view; but there is no doubt that in this way the circularity of impredicative definitions disappears.

Secondly, however, even if "all" means an infinite conjunction, it seems that the vicious circle principle in its first form applies only if the entities involved are constructed by ourselves. In this case there must clearly exist a definition (namely the description of the construction) which does not refer to a totality to which the object defined belongs, because the construction of a thing can certainly not be based on a totality of things to which the thing to be constructed itself belongs. If, however, it is a question of objects that exist independently of our constructions, there is nothing in the least absurd in the existence of totalities containing members, which can be described (i.e., uniquely characterized)[14] only by reference to this totality (cf. Ramsey 1926a: 338 or 1931: 1). Such a state of affairs would not even contradict the second form of the vicious circle principle, since one cannot say that an object described by reference to a totality "involves" this totality, although the description itself does; nor would it contradict the third form, if "presuppose" means "presuppose for the existence" not "for the knowability."

So it seems that the vicious circle principle in its first form applies only if one takes the constructivistic (or nominalistic) standpoint[15] toward the objects of logic and mathematics, in particular toward propositions, classes and notions, e.g., if one understands by a notion a symbol together with a rule for translating sentences containing the symbol into such sentences as do not contain it, so that a separate object denoted by the symbol appears as a mere fiction.[16]

Classes and concepts may, however, also be conceived as real objects, namely classes as "pluralities of things" or as structures consisting of a plurality of things and concepts as the properties and relations of things existing independently of our definitions and constructions.

It seems to me that the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a

[14]An object a is said to be described by a propositional function $\varphi(x)$ is $\varphi(x)$ is true for $x=a$ and for no other object.

[15]I shall use in the sequel "constructivism" as a general term comprising both these standpoints and also such tendencies as are embodied in Russell's "no class" theory.

[16]One might think that this conception of notions is impossible, because the sentences into which one translates must also contain notions so that one would get into an infinite regress. This, however, does not preclude the possibility of maintaining the above viewpoint for all the more abstract notions, such as those of the second and higher types, or in fact for all notions except the primitive terms which might be only a very few.

satisfactory theory of our sense perceptions and in both cases it is impossible to interpret the propositions one wants to assert about these entities as propositions about the "data," i.e., in the latter case the actually occurring sense perceptions. Russell himself concludes in the last chapter of his book on *Meaning and Truth* (1940), though "with hesitation," that there exist "universals," but apparently he wants to confine this statement to concepts of sense perceptions, which does not help the logician. I shall use the term "concept" in the sequel exclusively in this objective sense. One formal difference between the two conceptions of notions would be that any two different definitions of the form $\alpha(x)=\varphi(x)$ can be assumed to define two different notions α in the constructivistic sense. (In particular this would be the case for the nominalistic interpretation of the term "notion" suggested above, since two such definitions give different rules of translation for propositions containing α.) For concepts, on the contrary, this is by no means the case, since the same thing may be described in different ways. It might even be that the axiom of extensionality[17] or at least something near to it holds for concepts. The difference may be illustrated by the following definition of the number two: "Two is the notion under which fall all pairs and nothing else." There is certainly more than one notion in the constructivistic sense satisfying this condition, but there might be one common "form" or "nature" of all pairs.

Since the vicious circle principle, in its first form does apply to constructed entities, impredicative definitions and the totality of all notions or classes or propositions are inadmissible in constructivistic logic. What an impredicative definition would require is to construct a notion by a combination of a set of notions to which the notion to be formed itself belongs. Hence if one tries to effect a retranslation of a sentence containing a symbol for such an impredicatively defined notion it turns out that what one obtains will again contain a symbol for the notion in question (cf. Carnap 1931: 103 [51 in this volume] and 1937: 162). At least this is so if "all" means an infinite conjunction; but Carnap's and Langford's idea (mentioned on pp. [455–6]) would not help in this connection, because "demonstrability," if introduced in a manner compatible with the constructivistic standpoint towards notions, would have to be split into a hierarchy of orders, which would prevent one from obtaining the desired results.[18] As Chwistek (1933: 367) has shown, it is even possible under

[17]I.e., that no two different properties belong to exactly the same things, which, in a sense, is a counterpart to Leibniz's *Principium identitatis indiscernibilium*, which says no two different things have exactly the same properties.

[18]Nevertheless the scheme is interesting because it again shows the constructibility of notions which can be meaningfully asserted of notions of arbitrarily high order.

certain assumptions admissible within constructivistic logic to derive an actual contradiction from the unrestricted admission of impredicative definitions. To be more specific, he has shown that the system of simple types becomes contradictory if one adds the "axiom of intensionality" which says (roughly speaking) that to different definitions belong different notions. This axiom, however, as has just been pointed out, can be assumed to hold for notions in the constructivistic sense.

Speaking of concepts, the aspect of the question is changed completely. Since concepts are supposed to exist objectively, there seems to be objection neither to speaking of all of them (cf. p. [461]) nor to describing some of them by reference to all (or at least all of a given type). But, one may ask, isn't this view refutable also for concepts because it leads to the "absurdity" that there will exist properties φ such that $\varphi(a)$ consists in a certain state of affairs involving all properties (including φ itself and properties defined in terms of φ), which would mean that the vicious circle principle does not hold even in its second form for concepts or propositions? There is no doubt that the totality of all properties (or of all those of a given type) does lead to situations of this kind, but I don't think they contain any absurdity.[19] It is true that such properties φ [or such propositions $\varphi(a)$] will have to contain themselves as constituents of their content [or of their meaning], and in fact in many ways, because of the properties defined in terms of φ; but this only makes it impossible to construct their meaning (i.e., explain it as an assertion about sense perceptions or any other non-conceptual entities), which is no objection for one who takes the realistic standpoint. Nor is it self-contradictory that a proper part should be identical (not merely equal) to the whole, as is seen in the case of structures in the abstract sense. The structure of the series of integers, e.g., contains itself as a proper part and it is easily seen that there exist also structures containing infinitely many different parts, each containing the whole structure as a part. In addition there exist, even within the domain of constructivistic logic, certain approximations to this self-reflexivity of impredicative properties, namely propositions which contain as parts of their meaning not themselves but their own formal demonstrability (cf. Gödel 1931: 173 or Carnap 1937, §35). Now formal demonstrability of a proposition (in case the axioms and rules of inference are correct) implies this proposition

[19]The formal system corresponding to this view would have, instead of the axiom of reducibility, the rule of substitution for functions described, e.g., in Hilbert-Bernays 1934–9, 1: 90, applied to variables of any type, together with certain axioms of intensionality required by the concept of property which, however, would be weaker than Chwistek's. It should be noted that this view does not necessarily imply the existence of concepts which cannot be expressed in the system, if combined with a solution of the paradoxes along the lines indicated on p. [466].

and in many cases is equivalent to it. Furthermore, there doubtlessly exist sentences referring to a totality of sentences to which they themselves belong as, e.g., the sentence: "Every sentence (of a given language) contains at least one relation word."

Of course this view concerning the impredicative properties makes it necessary to look for another solution of the paradoxes, according to which the fallacy (i.e., the underlying erroneous axiom) does not consist in the assumption of certain self-reflexivities of the primitive terms but in other assumptions about these. Such a solution may be found for the present in the simple theory of types and in the future perhaps in the development of the ideas sketched on pp. [452–3 and 466]. Of course, all this refers only to concepts. As to notions in the constructivistic sense there is no doubt that the paradoxes are due to a vicious circle. It is not surprising that the paradoxes should have different solutions for different interpretations of the terms occurring.

As to classes in the sense of pluralities or totalities it would seem that they are likewise not created but merely described by their definitions and that therefore the vicious circle principle in the first form does not apply. I even think there exist interpretations of the term "class" (namely as a certain kind of structures), where it does not apply in the second form either.[20] But for the development of all contemporary mathematics one may even assume that it does apply in the second form, which for classes as mere pluralities is, indeed, a very plausible assumption. One is then led to something like Zermelo's axiom system for set theory, i.e., the sets are split up into "levels" in such a manner that only sets of lower levels can be elements of sets of higher levels (i.e., $x \in y$ is always false if x belongs to a higher level than y). There is no reason for classes in this sense to exclude mixtures of levels in one set and transfinite levels. The place of the axiom of reducibility is now taken by the axiom of classes [Zermelo's *Aussonderungsaxiom*] which says that for each level there exists for an arbitrary propositional function $\varphi(x)$ the set of those x of this level for which $\varphi(x)$ is true, and this seems to be implied by the concept of classes as pluralities.

Russell adduces two reasons against the extensional view of classes, namely the existence of (1) the null class, which cannot very well be a collection, and (2) the unit classes, which would have to be identical with their single elements. But it seems to me that these arguments could, if anything, at most prove that the null class and the unit classes (as distinct from their only element) are fictions (introduced to simplify the calculus like the points at infinity in geometry), not that all classes are fictions.

[20]Ideas tending in this direction are contained in Mirimanoff 1917a: 37–52, 1917b: 209–17, 1920: 29–52.

But in Russell the paradoxes had produced a pronounced tendency to build up logic as far as possible without the assumption of the objective existence of such entities as classes and concepts. This led to the formulation of the aforementioned "no class theory," according to which classes and concepts were to be introduced as a *façon de parler*. But propositions, too, (in particular those involving quantifications; Russell 1906a: 627) were later on largely included in this scheme, which is but a logical consequence of this standpoint, since, e.g., universal propositions as objectively existing entities evidently belong to the same category of idealistic objects as classes and concepts and lead to the same kind of paradoxes, if admitted without restrictions. As regards classes this program was actually carried out, i.e., the rules for translating sentences containing class names or the term "class" into such as do not contain them were stated explicitly; and the basis of the theory, i.e., the domain of sentences into which one has to translate is clear, so that classes can be dispensed with (within the system *Principia*), but only if one assumes the existence of a concept whenever one wants to construct a class. When it comes to concepts and the interpretation of sentences containing this or some synonymous term, the state of affairs is by no means as clear. First of all, some of them (the primitive predicates and relations such as "red" or "colder") must apparently be considered as real objects;[21] the rest of them (in particular according to the second edition of *Principia*, all notions of a type higher than the first and therewith all logically interesting ones) appear as something constructed (i.e., as something not belonging to the "inventory" of the world); but neither the basic domain of propositions in terms of which finally everything is to be interpreted, nor the method of interpretation is as clear as in the case of classes (see below).

This whole scheme of the no-class theory is of great interest as one of the few examples, carried out in detail, of the tendency to eliminate assumptions about the existence of objects outside the "data" and to replace them by constructions on the basis of these data.[22] The result has been in this case essentially negative; i.e., the classes and concepts introduced in this way do not have all the properties required for their use in mathematics, unless one either introduces special axioms about the data (e.g., the axiom of reducibility), which in essence already mean the existence in the data of the kind of objects to be constructed, or makes the fiction that one can form propositions of infinite (and even non-

[21] In Appendix C of *Principia* a way is sketched by which these also could be constructed by means of certain similarity relations between atomic propositions, so that these latter would be the only ones remaining as real objects.

[22] The "data" are to be understood in a relative sense here, i.e., in our case as logic without the assumption of the existence of classes and concepts.

denumerable) length (cf. Ramsey 1926a: 338 or 1931: 1), i.e., operates with truth-functions of infinitely many arguments, regardless of whether or not one can construct them. But what else is such an infinite truth-function but a special kind of an infinite extension (or structure) and even a more complicated one than a class, endowed in addition with a hypothetical meaning, which can be understood only by an infinite mind? All this is only a verification of the view defended above that logic and mathematics (just as physics) are built up on axioms with a real content which cannot be "explained away."

What one can obtain on the basis of the constructivistic attitude is the theory of orders (cf. p. [454]); only now (and this is the strong point of the theory) the restrictions involved do not appear as *ad hoc* hypotheses for avoiding the paradoxes, but as unavoidable consequences of the thesis that classes, concepts, and quantified propositions do not exist as real objects. It is not as if the universe of things were divided into orders and then one were prohibited to speak of all orders; but, on the contrary, it is possible to speak of all existing things; only, classes and concepts are not among them; and if they are introduced as a *façon de parler*, it turns out that this very extension of the symbolism gives rise to the possibility of introducing them in a more comprehensive way, and so on indefinitely. In order to carry out this scheme one must, however, presuppose arithmetic (or something equivalent) which only proves that not even this restricted logic can be built up on nothing.

In the first edition of *Principia*, where it was a question of actually building up logic and mathematics, the constructivistic attitude was, for the most part, abandoned, since the axiom of reducibility for types higher than the first together with the axiom of infinity makes it absolutely necessary that there exist primitive predicates of arbitrarily high types. What is left of the constructive attitude is only: (1) The introduction of classes as a *façon de parler;* (2) the definition of $\sim, \vee, \cdot$, etc., as applied to propositions containing quantifiers (which incidentally proved its fecundity in a consistency proof for arithmetic); (3) the step-by-step construction of functions of orders higher than 1, which, however, is superfluous owing to the axiom of reducibility; (4) the interpretation of definitions as mere typographical abbreviations, which makes every symbol introduced by definition an incomplete symbol (not one naming an object described by the definition). But the last item is largely an illusion, because, owing to the axiom of reducibility, there always exist real objects in the form of primitive predicates, or combinations of such, corresponding to each defined symbol. Finally also Russell's theory of descriptions is something belonging to the constructivistic order of ideas.

In the second edition of *Principia* (or to be more exact, in the introduction to it) the constructivistic attitude is resumed again. The axiom of reducibility is dropped and it is stated explicitly that all primitive predicates belong to the lowest type and that the only purpose of variables (and evidently also of constants) of higher orders and types is to make it possible to assert more complicated truth-functions of atomic propositions,[23] which is only another way of saying that the higher types and orders are solely a *façon de parler*. This statement at the same time informs us of what kind of propositions the basis of the theory is to consist, namely of truth-functions of atomic propositions.

This, however, is without difficulty only if the number of individuals and primitive predicates is finite. For the opposite case (which is chiefly of interest for the purpose of deriving mathematics), Ramsey (cf. Ramsey 1926a: 338 or 1931: 1) took the course of considering our inability to form propositions of infinite length as a "mere accident," to be neglected by the logician. This of course solves (or rather cuts through) the difficulties; but it is to be noted that, if one disregards the difference between finite and infinite in this respect, there exists a simpler and at the same time more far-reaching interpretation of set theory (and therewith of mathematics). Namely, in case of a finite number of individuals, Russell's *aperçu* that propositions about classes can be interpreted as propositions about their elements becomes literally true, since, e.g., "$x \in m$" is equivalent to "$x=a_1 \vee x=a_2 \vee \ldots \vee x=a_k$" where the a_i are the elements of m; and "there exists a class such that . . ." is equivalent to "there exist individuals $x_1, x_2, \ldots, x_n$ such that . . .,"[24] provided n is the number of individuals in the world and provided we neglect for the moment the null class which would have to be taken care of by an additional clause. Of course, by an iteration of this procedure one can obtain classes of classes, etc., so that the logical system obtained would resemble the theory of simple types except for the circumstance that mixture of types would be possible. Axiomatic set theory appears, then, as an extrapolation of this scheme for the case of infinitely many individuals or an infinite iteration of the process of forming sets.

Ramsey's viewpoint is, of course, everything but constructivistic, unless one means constructions of an infinite mind. Russell, in the second edition of *Principia,* took a less metaphysical course by confining himself to such truth-functions as can actually be constructed. In this way one is again led to the theory of orders, which, however, appears now in a new light, namely as a method of constructing more and more complicated

[23] I.e., propositions of the form $S(a), R(a,b)$, etc., where S, R are primitive predicates and a, b individuals.

[24] The x_i may, of course, as always, be partly or wholly identical with each other.

truth-functions of atomic propositions. But this procedure seems to presuppose arithmetic in some form or other (see next paragraph).

As to the question of how far mathematics can be built up on this basis (without any assumptions about the data – i.e., about the primitive predicates and individuals – except, as far as necessary, the axiom of infinity), it is clear that the theory of real numbers in its present form cannot be obtained.[25] As to the theory of integers, it is contended in the second edition of *Principia* that it can be obtained. The difficulty to be overcome is that in the definition of the integers as "those cardinals which belong to every class containing 0 and containing $x+1$ if containing x," the phrase "every class" must refer to a given order. So one obtains integers of different orders, and complete induction can be applied to integers of order n only for properties of order n; whereas it frequently happens that the notion of integer itself occurs in the property to which induction is applied. This notion, however, is of order $n+1$ for the integers of order n. Now, in Appendix B of the second edition of *Principia,* a proof is offered that the integers of any order higher than 5 are the same as those of order 5, which of course would settle all difficulties. The proof as it stands, however, is certainly not conclusive. In the proof of the main lemma *89.16, which says that every subset α (of arbitrary high order)[26] of an inductive class β of order 3 is itself an inductive class of order 3, induction is applied to a property of β involving α [namely $\alpha-\beta \neq \Lambda$, which, however, should read $\alpha - \beta \sim \in \mathrm{Induct}_2$ because (3) is evidently false]. This property, however, is of an order >3 if α is of an order >3. So the question whether (or to what extent) the theory of integers can be obtained on the basis of the ramified hierarchy must be considered as unsolved at the present time. It is to be noted, however, that, even in case this question should have a positive answer, this would be of no value for the problem whether arithmetic follows from logic, if propositional functions of order n are defined (as in the second edition of *Principia*) to be certain finite (though arbitrarily complex) combinations (of quantifiers, propositional connectives, etc.), because then the notion of finiteness has to be presupposed, which fact is concealed only by taking such complicated notions as "propositional function of order n" in an unanalyzed form as primitive terms of the formalism and giving their definition only in ordinary language. The reply may perhaps be

[25] As to the question how far it is possible to build up the theory of real numbers, presupposing the integers, cf. Weyl 1918.

[26] That the variable α is intended to be of undetermined order is seen from the later applications of *89.17 and from the note to *89.17. The main application is in line (2) of the proof of *89.24, where the lemma under consideration is needed for α's of arbitrarily high orders.

offered that in *Principia* the notion of a propositional function of order n is neither taken as primitive nor defined in terms of the notion of a finite combination, but rather quantifiers referring to propositional functions of order n (which is all one needs) are defined as certain infinite conjunctions and disjunctions. But then one must ask: Why doesn't one define the integers by the infinite disjunction: $x=0 \vee x=0+1 \vee x=0+1+1 \vee \ldots$ *ad infinitum*, saving in this way all the trouble connected with the notion of inductiveness? This whole objection would not apply if one understands by a propositional function of order n one "obtainable from such truth-functions of atomic propositions as presuppose for their definition no totalities except those of the propositional functions of order $<n$ and of individuals"; this notion, however, is somewhat lacking in precision.

The theory of orders proves more fruitful if considered from a purely mathematical standpoint, independently of the philosophical question whether impredicative definitions are admissible. Viewed in this manner, i.e., as a theory built up within the framework of ordinary mathematics, where impredicative definitions are admitted, there is no objection to extending it to arbitrarily high transfinite orders. Even if one rejects impredicative definitions, there would, I think, be no objection to extend it to such transfinite ordinals as can be constructed within the framework of finite orders. The theory in itself seems to demand such an extension since it leads automatically to the consideration of functions in whose definition one refers to all functions of finite orders, and these would be functions of order ω. Admitting transfinite orders, an axiom of reducibility can be proved. This, however, offers no help to the original purpose of the theory, because the ordinal α – such that every propositional function is extensionally equivalent to a function of order α – is so great, that it presupposes impredicative totalities. Nevertheless, so much can be accomplished in this way, that all impredicativities are reduced to one special kind, namely the existence of certain large ordinal numbers (or, well-ordered sets) and the validity of recursive reasoning for them. In particular, the existence of a well-ordered set, of order type ω_1 already suffices for the theory of real numbers. In addition this transfinite theorem of reducibility permits the proof of the consistency of the Axiom of Choice, of Cantor's Continuum-Hypothesis and even of the generalized Continuum-Hypothesis (which says that there exists no cardinal number between the power of any arbitrary set and the power of the set of its subsets) with the axioms of set theory as well as of *Principia*.

I now come in somewhat more detail to the theory of simple types which appears in *Principia* as combined with the theory of orders; the former is, however, (as remarked above) quite independent of the latter,

since mixed types evidently do not contradict the vicious circle principle in any way. Accordingly, Russell also based the theory of simple types on entirely different reasons. The reason adduced (in addition to its "consonance with common sense") is very similar to Frege's, who, in his system, already had assumed the theory of simple types for functions, but failed to avoid the paradoxes, because he operated with classes (or rather functions in extension) without any restriction. This reason is that (owing to the variable it contains) a propositional function is something ambiguous (or, as Frege says, something unsaturated, wanting supplementation) and therefore can occur in a meaningful proposition only in such a way that this ambiguity is eliminated (e.g., by substituting a constant for the variable or applying quantification to it). The consequences are that a function cannot replace an individual in a proposition, because the latter has no ambiguity to be removed, and that functions with different kinds of arguments (i.e., different ambiguities) cannot replace each other; which is the essence of the theory of simple types. Taking a more nominalistic viewpoint (such as suggested in the second edition of *Principia* and in *Meaning and Truth*) one would have to replace "proposition" by "sentence" in the foregoing considerations (with corresponding additional changes). But in both cases, this argument clearly belongs to the order of ideas of the "no class" theory, since it considers the notions (or propositional functions) as something constructed out of propositions or sentences by leaving one or several constituents of them undetermined. Propositional functions in this sense are so to speak "fragments" of propositions, which have no meaning in themselves, but only insofar as one can use them for forming propositions by combining several of them, which is possible only if they "fit together," i.e., if they are of appropriate types. But, it should be noted that the theory of simple types (in contradistinction to the vicious circle principle) cannot in a strict sense follow from the constructive standpoint, because one might construct notions and classes in another way, e.g., as indicated on p. [462], where mixtures of types are possible. If on the other hand one considers concepts as real objects, the theory of simple types is not very plausible, since what one would expect to be a concept (such as, e.g., "transitivity" or the number two) would seem to be something behind all its various "realizations" on the different levels and therefore does not exist according to the theory of types. Nevertheless, there seems to be some truth behind this idea of realizations of the same concept on various levels, and one might, therefore, expect the theory of simple types to prove useful or necessary at least as a stepping-stone for a more satisfactory system, a way in which it has already been used by Quine (cf. 1937: 70). Also

Russell's "typical ambiguity" is a step in this direction. Since, however, it only adds certain simplifying symbolic conventions to the theory of types, it does not *de facto* go beyond this theory.

It should be noted that the theory of types brings in a new idea for the solution of the paradoxes, especially suited to their intensional form. It consists in blaming the paradoxes not on the axiom that every propositional function defines a concept or class, but on the assumption that every concept gives a meaningful proposition, if asserted for any arbitrary object or objects as arguments. The obvious objection that every concept can be extended to all arguments, by defining another one which gives a false proposition whenever the original one was meaningless, can easily be dealt with by pointing out that the concept "meaningfully applicable" need not itself be always meaningfully applicable.

The theory of simple types (in its realistic interpretation) can be considered as a carrying through of this scheme, based, however, on the following additional assumption concerning meaningfulness: "Whenever an object x can replace another object y in one meaningful proposition, it can do so in every meaningful proposition."[27] This of course has the consequence that the objects are divided into mutually exclusive ranges of significance, each range consisting of those objects which can replace each other; and that therefore each concept is significant only for arguments belonging to one of these ranges, i.e., for an infinitely small portion of all objects. What makes the above principle particularly suspect, however, is that its very assumption makes its formulation as a meaningful proposition impossible,[28] because x and y must then be confined to definite ranges of significance which are either the same or different, and in both cases the statement does not express the principle or even part of it. Another consequence is that the fact that an object x is (or is not) of a given type also cannot be expressed by a meaningful proposition.

It is not impossible that the idea of limited ranges of significance could be carried out without the above restrictive principle. It might even turn out that it is possible to assume every concept to be significant everywhere except for certain "singular points" or "limiting points," so that the paradoxes would appear as something analogous to dividing by zero. Such a system would be most satisfactory in the following respect: our logical intuitions would then remain correct up to certain minor corrections, i.e., they could then be considered to give an essentially correct, only somewhat "blurred," picture of the real state of affairs. Unfortunately the attempts made in this direction have failed so far;[29] on the other hand, the impossibility of this scheme has not been proved either, in spite of the strong inconsistency theorems of Kleene and Rosser (1935: 630).

[27] Russell formulates a somewhat different principle with the same effect (Whitehead and Russell 1910–13, 1: 95).

[28] This objection does not apply to the symbolic interpretation of the theory of types, spoken of on p. [465], because there one does not have objects but only symbols of different types.

In conclusion I want to say a few words about the question whether (and in which sense) the axioms of *Principia* can be considered to be analytic. As to this problem it is to be remarked that analyticity may be understood in two senses. First, it may have the purely formal sense that the terms occurring can be defined (either explicitly or by rules for eliminating them from sentences containing them) in such a way that the axioms and theorems become special cases of the law of identity and disprovable propositions become negations of this law. In this sense even the theory of integers is demonstrably non-analytic, provided that one requires of the rules of elimination that they allow one actually to carry out the elimination in a finite number of steps in each case.[30] Leaving out this condition by admitting, e.g., sentences of infinite (and non-denumerable) length as intermediate steps of the process of reduction, all axioms of *Principia* (including the axioms of choice, infinity and reducibility) could be proved to be analytic for certain interpretations (by considerations similar to those referred to on p. [462]).[31] But this observation is of doubtful value, because the whole of mathematics as applied to sentences of infinite length has to be presupposed in order to prove this analyticity, e.g., the axiom of choice can be proved to be analytic only if it is assumed to be true.

In a second sense a proposition is called analytic if it holds, "owing to the meaning of the concepts occurring in it," where this meaning may perhaps be undefinable (i.e., irreducible to anything more fundamental).[32] It would seem that all axioms of *Principia*, in the first edition, (except the axiom of infinity) are in this sense analytic for certain interpretations of the primitive terms, namely if the term "predicative function" is replaced either by "class" (in the extensional sense) or (leaving out the axiom of choice) by "concept," since nothing can express better the

[29] A formal system along these lines is Church's "A Set of Postulates for the Foundation of Logic" (1932: 346; 1933: 839), where, however, the underlying idea is expressed by the somewhat misleading statement that the law of excluded middle is abandoned. However, this system has been proved to be inconsistent. See Kleene and Rosser 1935.

[30] Because this would imply the existence of a decision-procedure for all arithmetical propositions. Cf. Turing 1937: 230.

[31] Cf. also Ramsey (1926a: 338 or 1931: 1), where, however, the axiom of infinity cannot be obtained, because it is interpreted to refer to the individuals in the world.

[32] The two significations of the term *analytic* might perhaps be distinguished as tautological and analytic.

meaning of the term "class" than the axiom of the classes (cf. p. [459]) and the axiom of choice, and since, on the other hand, the meaning of the term "concept" seems to imply that every propositional function defines a concept.[33] The difficulty is only that we don't perceive the concepts of "concept" and of "class" with sufficient distinctness, as is shown by the paradoxes. In view of this situation, Russell took the course of considering both classes and concepts (except the logically uninteresting primitive predicates) as non-existent and of replacing them by constructions of our own. It cannot be denied that this procedure has led to interesting ideas and to results valuable also for one taking the opposite viewpoint. On the whole, however, the outcome has been that only fragments of Mathematical Logic remain, unless the things condemned are reintroduced in the form of infinite propositions or by such axioms as the axiom of reducibility which (in case of infinitely many individuals) is demonstrably false unless one assumes either the existence of classes or of infinitely many "*qualitates occultae.*" This seems to be an indication that one should take a more conservative course, such as would consist in trying to make the meaning of the terms "class" and "concept" clearer, and to set up a consistent theory of classes and concepts as objectively existing entities. This is the course which the actual development of Mathematical Logic has been taking and which Russell himself has been forced to enter upon in the more constructive parts of his work. Major among the attempts in this direction (some of which have been quoted in this essay) are the simple theory of types (which is the system of the first edition of *Principia* in an appropriate interpretation) and axiomatic set theory, both of which have been successful at least to this extent, that they permit the derivation of modern mathematics and at the same time avoid all known paradoxes. Many symptoms show only too clearly, however, that the primitive concepts need further elucidation.

It seems reasonable to suspect that it is this incomplete understanding of the foundations which is responsible for the fact that Mathematical Logic has up to now remained so far behind the high expectations of Peano and others who (in accordance with Leibniz's claims) had hoped

[33]This view does not contradict the opinion defended above that mathematics is based on axioms with a real content, because the very existence of the concept of, e.g., "class" constitutes already such an axiom; since, if one defined, e.g., "class" and "$\in$" to be "the concepts satisfying the axioms," one would be unable to prove their existence. "Concept" could perhaps be defined in terms of "proposition" (cf. p. [465]) (although I don't think that this would be a natural procedure); but then certain axioms about propositions, justifiable only with reference to the undefined meaning of this term, will have to be assumed. It is to be noted that this view about analyticity makes it again possible that every mathematical proposition could perhaps be reduced to a special case of $a = a$, namely if the reduction is effected not in virtue of the definitions of the terms occurring, but in virtue of their meaning, which can never be completely expressed in a set of formal rules.

that it would facilitate theoretical mathematics to the same extent as the decimal system of numbers has facilitated numerical computations. For how can one expect to solve mathematical problems systematically by mere analysis of the concepts occurring, if our analysis so far does not even suffice to set up the axioms? But there is no need to give up hope. Leibniz did not in his writings about the *Characteristica universalis* speak of a utopian project; if we are to believe his words he had developed this calculus of reasoning to a large extent, but was waiting with its publication till the seed could fall on fertile ground (1875–90, 7: 12; Vacca 1903: 72; Leibniz 1923–, 1: preface). He went even so far (1875–90, 7: 187) as to estimate the time which would be necessary for his calculus to be developed by a few select scientists to such an extent "that humanity would have a new kind of an instrument increasing the powers of reason far more than any optical instrument has ever aided the power of vision." The time he names is five years, and he claims that his method is not any more difficult to learn than the mathematics or philosophy of his time. Furthermore, he said repeatedly that, even in the rudimentary state to which he had developed the theory himself, it was responsible for all his mathematical discoveries; which, one should expect, even Poincaré would acknowledge as a sufficient proof of its fecundity.[34]

[34]I wish to express my thanks to Professor Alonzo Church, of Princeton University, who helped me find the correct English expressions in a number of places.

What is Cantor's continuum problem?

KURT GÖDEL

1. The concept of cardinal number

Cantor's continuum problem is simply the question: How many points are there on a straight line in euclidean space? An equivalent question is: How many different sets of integers do there exist?

This question, of course, could arise only after the concept of "number" had been extended to infinite sets; hence it might be doubted if this extension can be effected in a uniquely determined manner and if, therefore, the statement of the problem in the simple terms used above is justified. Closer examination, however, shows that Cantor's definition of infinite numbers really has this character of uniqueness. For whatever "number" as applied to infinite sets may mean, we certainly want it to have the property that the number of objects belonging to some class does not change if, leaving the objects the same, one changes in any way whatsoever their properties or mutual relations (e.g., their colors or their distribution in space). From this, however, it follows at once that two sets (at least two sets of changeable objects of the space-time world) will have the same cardinal number if their elements can be brought into a one-to-one correspondence, which is Cantor's definition of equality between numbers. For if there exists such a correspondence for two sets A and B it is possible (at least theoretically) to change the properties and relations of each element of A into those of the corresponding element of B, whereby A is transformed into a set completely indistinguishable from B, hence of the same cardinal number. For example, assuming a square and a line segment both completely filled with mass points (so that at each point of them exactly one mass point is situated), it follows, owing to the demonstrable fact that there exists a one-to-one correspondence between the points of a square and of a line segment and, therefore, also between the corresponding mass points, that the mass points of the square can be so rearranged as exactly to fill out the line segment, and

This is a revised and expanded version of a paper of the same title that appeared in *The American Mathematical Monthly* 54 (1947): 515–25. It is printed here with the kind permission of The Mathematical Association of America.

vice versa. Such considerations, it is true, apply directly only to physical objects, but a definition of the concept of "number" which would depend on the kind of objects that are numbered could hardly be considered to be satisfactory.

So there is hardly any choice left but to accept Cantor's definition of equality between numbers, which can easily be extended to a definition of "greater" and "less" for infinite numbers by stipulating that the cardinal number M of a set A is to be called less than the cardinal number N of a set B if M is different from N but equal to the cardinal number of some subset of B. That a cardinal number having a certain property exists is defined to mean that a set of such a cardinal number exists. On the basis of these definitions, it becomes possible to prove that there exist infinitely many different infinite cardinal numbers or "powers," and that, in particular, the number of subsets of a set is always greater than the number of its elements; furthermore, it becomes possible to extend (again without any arbitrariness) the arithmetical operations to infinite numbers (including sums and products with any infinite number of terms or factors) and to prove practically all ordinary rules of computation.

But, even after that, the problem of identifying the cardinal number of an individual set, such as the linear continuum, would not be well-defined if there did not exist some systematic representation of the infinite cardinal numbers, comparable to the decimal notation of the integers. Such a systematic representation, however, does exist, owing to the theorem that for each cardinal number and each set of cardinal numbers[1] there exists exactly one cardinal number immediately succeeding in magnitude and that the cardinal number of every set occurs in the series thus obtained.[2] This theorem makes it possible to denote the cardinal number immediately succeeding the set of finite numbers by $\aleph_0$ (which is the power of the "denumerably infinite" sets), the next one by $\aleph_1$, etc.; the one immediately succeeding all $\aleph_i$ where i is an integer, by $\aleph_\omega$, the next one by $\aleph_{\omega+1}$, etc. The theory of ordinal numbers provides the means for extending this series further and further.

[1]As to the question of why there does not exist a set of all cardinal numbers, see footnote 12.

[2]The axiom of choice is needed for the proof of this theorem (see Fraenkel and Bar-Hillel 1958). But it may be said that this axiom, from almost every possible point of view, is as well-founded today as the other axioms of set theory. It has been proved consistent with the other axioms of set theory which are usually assumed, provided that these other axioms are consistent (see Gödel 1940). Moreover, it is possible to define in terms of any system of objects satisfying the other axioms a system of objects satisfying those axioms *and* the axiom of choice. Finally, the axiom of choice is just as evident as the other set-theoretical axioms for the "pure" concept of set explained in footnote 11.

2. The continuum problem, the continuum hypothesis, and the partial results concerning its truth obtained so far

So the analysis of the phrase "how many" unambiguously leads to a definite meaning for the question stated in the second line of this paper: The problem is to find out which one of the $\aleph$'s is the number of points of a straight line or (which is the same) of any other continuum (of any number of dimensions) in a euclidean space. Cantor, after having proved that this number is greater than $\aleph_0$, conjectured that it is $\aleph_1$. An equivalent proposition is this: Any infinite subset of the continuum has the power either of the set of integers or of the whole continuum. This is Cantor's continuum hypothesis.

But, although Cantor's set theory now has had a development of more than seventy years and the problem evidently is of great importance for it, nothing has been proved so far about the question what the power of the continuum is or whether its subsets satisfy the condition just stated, except (1) that the power of the continuum is not a cardinal number of a certain special kind, namely, not a limit of denumerably many smaller cardinal numbers,[3] and (2) that the proposition just mentioned about the subsets of the continuum is true for a certain infinitesimal fraction of these subsets, the analytic[4] sets.[5] Not even an upper bound, however large, can be assigned for the power of the continuum. Nor is the quality of the cardinal number of the continuum known any better than its quantity. It is undecided whether this number is regular or singular, accessible or inaccessible, and (except for König's negative result) what its character of confinality (see footnote 4) is. The only thing that is known, in addition to the results just mentioned, is a great number of consequences of, and some propositions equivalent to, Cantor's conjecture (Sierpiński 1934a).

This pronounced failure becomes still more striking if the problem is considered in its connection with general questions of cardinal arithmetic. It is easily proved that the power of the continuum is equal to $2^{\aleph_0}$. So the continuum problem turns out to be a question from the "multiplication table" of cardinal numbers, namely, the problem of evaluating a certain infinite product (in fact the simplest non-trivial one that can be formed). There is, however, not one infinite product (of factors >1) for

[3]See Hausdorff, *Mengenlehre* 1914: 68, or Bachmann 1955: 167. The discoverer of this theorem, J. König, asserted more than he had actually proved (1905: 177).

[4]See the list of definitions pp. 480–1.

[5]See Hausdorff 1914, 3rd ed.: 32. Even for complements of analytic sets the question is undecided at present, and it can be proved only that they either have the power $\aleph_0$ or $\aleph_1$ or continuum or are finite (see Kuratowski 1933–50, 1: 246).

which so much as an upper bound for its value can be assigned. All one knows about the evaluation of infinite products are two lower bounds due to Cantor and König (the latter of which implies the aforementioned negative theorem on the power of the continuum), and some theorems concerning the reduction of products with different factors to exponentiations and of exponentiations to exponentiations with smaller bases or exponents. These theorems reduce[6] the whole problem of computing infinite products to the evaluation of $\aleph_\alpha^{\mathrm{cf}(\aleph_\alpha)}$ and the performance of certain fundamental operations on ordinal numbers, such as determining the limit of a series of them. All products and powers, can easily be computed[7] if the "generalized continuum hypothesis" is assumed; i.e., if it is assumed that $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for every α, or, in other terms, that the number of subsets of a set of power $\aleph_\alpha$ is $\aleph_{\alpha+1}$. But, without making any undemonstrated assumption, it is not even known whether or not $m<n$ implies $2^m < 2^n$ (although it is trivial that it implies $2^m \leqslant 2^n$), nor even whether $2^{\aleph_0} < 2^{\aleph_1}$.

3. Restatement of the problem on the basis of an analysis of the foundations of set theory and results obtained along these lines

This scarcity of results, even as to the most fundamental questions in this field, to some extent may be due to purely mathematical difficulties; it seems, however (see Section 4), that there are also deeper reasons involved and that a complete solution of these problems can be obtained only by a more profound analysis (than mathematics is accustomed to giving) of the meanings of the terms occurring in them (such as "set", "one-to-one correspondence", etc.) and of the axioms underlying their use. Several such analyses have already been proposed. Let us see then what they give for our problem.

First of all there is Brouwer's intuitionism, which is utterly destructive in its results. The whole theory of the $\aleph$'s greater than $\aleph_1$ is rejected as meaningless (Brouwer 1909: 569). Cantor's conjecture itself receives several different meanings, all of which, though very interesting in themselves, are quite different from the original problem. They lead partly to affirmative, partly to negative answers (Brouwer 1907, I: 9; III: 2). Not everything in this field, however, has been sufficiently clarified. The

[6]This reduction can be effected, owing to the results and methods of a paper by Tarski (1925: 1).

[7]For regular numbers $\aleph_\alpha$, one obtains immediately:

$$\aleph_\alpha^{\mathrm{cf}(\aleph_\alpha)} = \aleph_\alpha^{\aleph_\alpha} = 2^{\aleph_\alpha} = \aleph_{\alpha+1}.$$

"semi-intuitionistic" standpoint along the lines of H. Poincaré and H. Weyl[8] would hardly preserve substantially more of set theory.

However, this negative attitude toward Cantor's set theory, and toward classical mathematics, of which it is a natural generalization, is by no means a necessary outcome of a closer examination of their foundations, but only the result of a certain philosophical conception of the nature of mathematics, which admits mathematical objects only to the extent to which they are interpretable as our own constructions or, at least, can be completely given in mathematical intuition. For someone who considers mathematical objects to exist independently of our constructions and of our having an intuition of them individually, and who requires only that the general mathematical concepts must be sufficiently clear for us to be able to recognize their soundness and the truth of the axioms concerning them, there exists, I believe, a satisfactory foundation of Cantor's set theory in its whole original extent and meaning, namely axiomatics of set theory interpreted in the way sketched below.

It might seem at first that the set-theoretical paradoxes would doom to failure such an undertaking, but closer examination shows that they cause no trouble at all. They are a very serious problem, not for mathematics, however, but rather for logic and epistemology. As far as sets occur in mathematics (at least in the mathematics of today, including all of Cantor's set theory), they are sets of integers, or of rational numbers (i.e., of pairs of integers), or of real numbers (i.e., of sets of rational numbers), or of functions of real numbers (i.e., of sets of pairs of real numbers), etc. When theorems about all sets (or the existence of sets in general) are asserted, they can always be interpreted without any difficulty to mean that they hold for sets of integers as well as for sets of sets of integers, etc. (respectively, that there either exist sets of integers, or sets of sets of integers, or ... etc., which have the asserted property). This concept of set,[9] however, according to which a set is something obtainable from the integers (or some other well-defined objects) by iterated

[8]See Weyl 1918 (1932 ed.). If the procedure of construction of sets described there (p. 20) is iterated, a sufficiently large (transfinite) number of times, one gets exactly the real numbers of the model for set theory mentioned in Section 4, in which the continuum hypothesis is true. But this iteration is not possible within the limits of the semi-intuitionistic standpoint.

[9]It must be admitted that the spirit of the modern abstract disciplines of mathematics, in particular of the theory of categories, transcends this concept of set, as becomes apparent, e.g., by the self-applicability of categories (see MacLane 1961). It does not seem, however, that anything is lost from the mathematical content of the theory if categories of different levels are distinguished. If there existed mathematically interesting proofs that would not go through under this interpretation, then the paradoxes of set theory would become a serious problem for mathematics.

application[10] of the operation "set of",[11] not something obtained by dividing the totality of all existing things into two categories, has never led to any antinomy whatsoever; that is, the perfectly "naïve" and uncritical working with this concept of set has so far proved completely self-consistent.[12]

But, furthermore, the axioms underlying the unrestricted use of this concept of set or, at least, a subset of them which suffices for all mathematical proofs devised up to now (except for theorems about the existence of extremely large cardinal numbers, see footnote 16), have been formulated so precisely in axiomatic set theory[13] that the question of whether some given proposition follows from them can be transformed, by means of mathematical logic, into a purely combinatorial problem concerning the manipulation of symbols which even the most radical intuitionist must acknowledge as meaningful. So Cantor's continuum problem, no matter what philosophical standpoint is taken, undeniably retains at least this meaning: to find out whether an answer, and if so which answer, can be derived from the axioms of set theory as formulated in the systems cited.

Of course, if it is interpreted in this way, there are (assuming the consistency of the axioms) *a priori* three possibilities for Cantor's conjecture: It may be demonstrable, disprovable, or undecidable.[14] The third alternative (which is only a precise formulation of the foregoing conjecture, that the difficulties of the problem are probably not purely mathematical), is the most likely. To seek a proof for it is, at present, perhaps the most promising way of attacking the problem. One result along these

[10]This phrase is meant to include transfinite iteration; i.e., the totality of sets obtained by finite iteration is considered to be itself a set and a basis for further applications of the operation "set of".

[11]The operation "set of *x*'s" (where the variable "*x*" ranges over some given kind of objects) cannot be defined satisfactorily (at least not in the present state of knowledge), but can only be paraphrased by other expressions involving again the concept of set, such as: "multitude" ("combination", "part") is conceived of as something which exists in itself no matter whether we can define it in a finite number of words (so that random sets are not excluded).

[12]It follows at once from this explanation of the term "set" that a set of all sets or other sets of a similar extension cannot exist, since every set obtained in this way immediately gives rise to further applications of the operation "set of" and, therefore, to the existence of larger sets.

[13]See, e.g., Bernays, 1937–54, 2: 65; 6: 1; 7: 65, 133; 8: 89. Von Neumann 1925: 219; cf. also von Neumann 1929: 227; and 1928: 669; Gödel 1940; Bernays 1958. By including very strong axioms of infinity, much more elegant axiomatizations have recently become possible. (See Bernays 1961.)

[14]In case the axioms were inconsistent the last one of the four *a priori* possible alternatives for Cantor's conjecture would occur, namely, it would then be both demonstrable and disprovable by the axioms of set theory.

lines has been obtained already, namely, that Cantor's conjecture is not disprovable from the axioms of set theory, provided that these axioms are consistent (see Section 4).

It is to be noted, however, that on the basis of the point of view here adopted, a proof of the undecidability of Cantor's conjecture from the accepted axioms of set theory (in contradistinction, e.g., to the proof of the transcendency of π) would by no means solve the problem. For if the meanings of the primitive terms of set theory as explained on pages [474–5] and in footnote 11 are accepted as sound, it follows that the set-theoretical concepts and theorems describe some well-determined reality, in which Cantor's conjecture must be either true or false. Hence its undecidability from the axioms being assumed today can only mean that these axioms do not contain a complete description of that reality. Such a belief is by no means chimerical, since it is possible to point out ways in which the decision of a question, which is undecidable from the usual axioms, might nevertheless be obtained.

First of all the axioms of set theory by no means form a system closed in itself, but, quite on the contrary, the very concept of set[15] on which they are based suggests their extension by new axioms which assert the existence of still further iterations of the operation "set of". These axioms can be formulated also as propositions asserting the existence of very great cardinal numbers (i.e., of sets having these cardinal numbers). The simplest of these strong "axioms of infinity" asserts the existence of inaccessible numbers (in the weaker or stronger sense) $>\aleph_0$. The latter axiom, roughly speaking, means nothing else but that the totality of sets obtainable by use of the procedures of formation of sets expressed in the other axioms forms again a set (and, therefore, a new basis for further applications of these procedures) (Zermelo 1930: 29). Other axioms of infinity have first been formulated by P. Mahlo.[16] These axioms show clearly, not only that the axiomatic system of set theory as used today is

[15]Similarly the concept "property of set" (the second of the primitive terms of set theory) suggests continued extensions of the axioms referring to it. Furthermore, concepts of "property of property of set" etc. can be introduced. The new axioms thus obtained, however, as to their consequences for propositions referring to limited domains of sets (such as the continuum hypothesis) are contained (as far as they are known today) in the axioms about sets.

[16]See Mahlo 1911: 190–200; 1913: 269–76. From Mahlo's presentation of the subject, however, it does not appear that the numbers he defines actually exist. In recent years considerable progress has been made as to the axioms of infinity. In particular, some have been formulated that are based on principles entirely different from those of Mahlo, and Dana Scott has proved that one of them implies the negation of proposition A (mentioned on p. [478]). So the consistency proof for the continuum hypothesis explained on p. [479] does *not* go through if this axiom is added. However, that these axioms are implied by the

incomplete, but also that it can be supplemented without arbitrariness by new axioms which only unfold the content of the concept of set explained above.

It can be proved that these axioms also have consequences far outside the domain of very great transfinite numbers, which is their immediate subject matter: each of them, under the assumption of its consistency, can be shown to increase the number of decidable propositions even in the field of Diophantine equations. As for the continuum problem, there is little hope of solving it by means of those axioms of infinity which can be set up on the basis of Mahlo's principles (the aforementioned proof for the undisprovability of the continuum hypothesis goes through for all of them without any change). But there exist others based on different principles (see footnote 16); also there may exist, besides the usual axioms, the axioms of infinity, and the axioms mentioned in footnote 15, other (hitherto unknown) axioms of set theory which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts (see, e.g., footnote 19).

Secondly, however, even disregarding the intrinsic necessity of some new axiom, and even in case it has no intrinsic necessity at all, a probable decision about its truth is possible also in another way, namely, inductively by studying its "success." Success here means fruitfulness in consequences, in particular in "verifiable" consequences, i.e., consequences demonstrable without the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover, and make it possible to contract into one proof many different proofs. The axioms for the system of real numbers, rejected by the intuitionists, have in this sense been verified to some extent, owing to the fact that analytical number theory frequently allows one to prove number-theoretical theorems which, in a more cumbersome way, can subsequently be verified by elementary methods. A much higher degree of verification than that, however, is conceivable. There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding such powerful methods for solving problems (and even solving them constructively, as far as that is possible) that, no matter whether or not they are intrinsically necessary, they would have to be accepted at least in the same sense as any well-established physical theory.

general concept of set in the same sense as Mahlo's has not been made clear yet. See Tarski 1962: 134; Scott 1961; Hanf and Scott 1961: 445. Mahlo's axioms have been derived by Azriel Lévy from a general principle about the system of all sets (1960: 233). See also Bernays 1961: 1, where almost all set-theoretical axioms are derived from Lévy's principle.

4. Some observations about the question: In what sense and in which direction may a solution of the continuum problem be expected?

But are such considerations appropriate for the continuum problem? Are there really any clear indications for its unsolvability by the accepted axioms? I think there are at least two:

The first results from the fact that there are two quite differently defined classes of objects both of which satisfy all axioms of set theory that have been set up so far. One class consists of the sets definable in a certain manner by properties of their elements;[17] the other of the sets in the sense of arbitrary multitudes, regardless of if, or how, they can be defined. Now, before it has been settled what objects are to be numbered, and on the basis of what one-to-one correspondences, one can hardly expect to be able to determine their number, expect perhaps in the case of some fortunate coincidence. If, however, one believes that it is meaningless to speak of sets except in the sense of extensions of definable properties, then, too, he can hardly expect more than a small fraction of the problems of set theory to be solvable without making use of this, in his opinion essential, characteristic of sets, namely, that they are extensions of definable properties. This characteristic of sets, however, is neither formulated explicitly nor contained implicitly in the accepted axioms of set theory. So from either point of view, if in addition one takes into account what was said in Section 2, it may be conjectured that the continuum problem cannot be solved on the basis of the axioms set up so far, but, on the other hand, may be solvable with the help of some new axiom which would state or imply something about the definability of sets.[18]

The latter half of this conjecture has already been verified; namely, the concept of definability mentioned in footnote 17 (which itself is definable in axiomatic set theory) makes it possible to derive, in axiomatic set theory, the generalized continuum hypothesis from the axiom that every set is definable in this sense.[19] Since this axiom (let us call it "A") turns out to be demonstrably consistent with the other axioms, under the

[17]Namely, definable by certain procedures, "*in terms* of ordinal numbers" (i.e., roughly speaking, under the assumption that for each ordinal number a symbol denoting it is given). See Gödel 1940 and 1939. The paradox of Richard, of course, does not apply to this kind of definability, since the totality of ordinals is certainly not denumerable.

[18]D. Hilbert's program for a solution of the continuum problem (see Hilbert 1926: 161), which, however, has never been carried through, also was based on a consideration of all possible definitions of real numbers.

[19]On the other hand, from an axiom in some sense opposite to this one, the negation of Cantor's conjecture could perhaps be derived. I am thinking of an axiom which (similar to Hilbert's completeness axiom in geometry) would state some maximum property of the

assumption of the consistency of these other axioms, this result (regardless of the philosophical position taken toward definability) shows the consistency of the continuum hypothesis with the axioms of set theory, provided that these axioms themselves are consistent.[20] The proof in its structure is similar to the consistency-proof of non-euclidean geometry by means of a model within euclidean geometry. Namely, it follows from the axioms of set theory that the sets definable in the aforementioned sense form a model of set theory in which the proposition A and, therefore, the generalized continuum hypothesis is true.

A second argument in favor of the unsolvability of the continuum problem on the basis of the usual axioms can be based on certain facts (not known at Cantor's time) which seem to indicate that Cantor's conjecture will turn out to be wrong,[21] while, on the other hand, a disproof of it is demonstrably impossible on the basis of the axioms being assumed today.

One such fact is the existence of certain properties of point sets (asserting an extreme rareness of the sets concerned) for which one has succeeded in proving the existence of non-denumerable sets having these properties, but no way is apparent in which one could expect to prove the existence of examples of the power of the continuum. Properties of this type (of subsets of a straight line) are: (1) being of the first category on every perfect set (Sierpiński 1934b: 270, and Kuratowski 1933–50, 1: 269ff), (2) being carried into a zero set by every continuous one-to-one mapping of the line onto itself (Lusin and Sierpiński 1918: 35, and Sierpiński 1934b: 270). Another property of a similar nature is that of being coverable by infinitely many intervals of any given lengths. But in this case one has so far not even succeeded in proving the existence of non-denumerable examples. From the continuum hypothesis, however, it follows in all three cases that there exist, not only examples of the power of the continuum,[22] but even such as are carried into themselves (up to denumerably many points) by *every* translation of the straight line (Sierpiński 1935a: 43).

Other highly implausible consequences of the continuum hypothesis are that there exist: (1) subsets of a straight line of the power of the con-

system of all sets, whereas axiom A states a minimum property. Note that only a maximum property would seem to harmonize with the concept of set explained in footnote 11.

[20]See Gödel 1940 and 1939: 220. I take this opportunity to correct a mistake in the notation and a misprint which occurred in this paper: in lines 25–29, p. 221; 4–6 and 10, p. 222; 11–19, p. 223, the letter α should be replaced (at all places where it occurs) by μ. Also, in Theorem 6, p. 222, the symbol "$\equiv$" should be inserted between $\phi_\alpha(x)$ and $\phi_{\bar{\alpha}}(x')$. For a carrying through of the proof in all details, Gödel 1940 is to be consulted.

[21]Views tending in this direction have been expressed also by Lusin 1935: 129ff. See also Sierpiński 1934a: 132.

[22]For the third case see Sierpiński 1934a (1st ed.): 39, Theorem 1.

tinuum which are covered (up to denumerably many points) by *every* dense set of intervals (Lusin 1914: 1259); (2) infinite dimensional subsets of Hilbert space which contain no non-denumerable finite-dimensional subset (in the sense of Menger-Urysohn) (Hurewicz 1932: 8); (3) an infinite sequence A^i of decompositions of any set M of the power of the continuum into continuum many mutually exclusive sets A^i_x such that, in whichever way a set $A^i_{x_i}$ is chosen for each i, $\prod_{i=0}^{\infty}(M - A^i_{x_i})$ is denumerable.[23] (1) and (3) are very implausible even if "power of the continuum" is replaced by "$\aleph_1$".

One may say that many results of point-set theory obtained without using the continuum hypothesis also are highly unexpected and implausible (Blumenthal 1940: 346). But, true as that may be, still the situation is different there, in that, in most of those instances (such as, e.g., Peano's curves), the appearance to the contrary can be explained by a lack of agreement between our intuitive geometrical concepts and the set-theoretical ones occurring in the theorems. Also, it is very suspicious that, as against the numerous plausible propositions which imply the negation of the continuum hypothesis, not one plausible proposition is known which would imply the continuum hypothesis. I believe that adding up all that has been said one has good reason for suspecting that the role of the continuum problem in set theory will be to lead to the discovery of new axioms which will make it possible to disprove Cantor's conjecture.

Definitions of some of the technical terms

Definitions 4–15 refer to subsets of a straight line, but can be literally transferred to subsets of euclidean spaces of any number of dimensions if "interval" is identified with "interior of a parallelepipedon."

1. I call *the character of confinality* of a cardinal number m (abbreviated by "cf(m)") the smallest number n such that m is the sum of n numbers $<m$.
2. A cardinal number m is *regular* if $\mathrm{cf}(m) = m$, otherwise singular.
3. An infinite cardinal number m is *inaccessible* if it is regular and has no immediate predecessor (i.e., if, although it is a limit of numbers $<m$, it is not a limit of fewer than m such numbers); it is *strongly inaccessible* if each product (and, therefore, also each sum) of fewer than m numbers $<m$ is $<m$. (See Sierpiński and Tarski 1930: 292; Tarski 1938: 68.)

[23] See Braun and Sierpiński 1932: 1, proposition (Q). This proposition is equivalent with the continuum hypothesis.

It follows from the generalized continuum hypothesis that these two concepts are equivalent. $\aleph_0$ is evidently inaccessible and also strongly inaccessible. As for finite numbers, 0 and 2 and no others are strongly inaccessible. A definition of inaccessibility, applicable to finite numbers, is this: m is inaccessible if (1) any sum of fewer than m numbers $<m$ is $<m$, and (2) the number of numbers $<m$ is m. This definition, for transfinite numbers, agrees with that given above, and for finite numbers yields $0, 1, 2$ as inaccessible. So inaccessibility and strong inaccessibility turn out not to be equivalent for finite numbers. This casts some doubt on their equivalence for transfinite numbers, which follows from the generalized continuum hypothesis.

4. A set of intervals is *dense* if every interval has points in common with some interval of the set. (The end-points of an interval are not considered as points of the interval.)
5. A *zero-set* is a set which can be covered by infinite sets of intervals with arbitrarily small lengths-sum.
6. A *neighborhood* of a point P is an interval containing P.
7. A subset A of B is *dense in B* if every neighborhood of any point of B contains points of A.
8. A point P is in the *exterior* of A if it has a neighborhood containing no point of A.
9. A subset A of B is *nowhere dense in B* if those points of B which are in the exterior of A are dense in B, or (which is equivalent) if for no interval I the intersection IA is dense in IB.
10. A subset A of B is *of the first category in B* if it is the sum of denumerably many sets nowhere dense in B.
11. A set A is *of the first category on B* if the intersection AB is of the first category in B.
12. A point P is called *a limit point* of a set A if any neighborhood of P contains infinitely many points of A.
13. A set A is called *closed* if it contains all its limit points.
14. A set is *perfect* if it is closed and has no isolated point (i.e., no point with a neighborhood containing no other point of the set).
15. *Borel-sets* are defined as the smallest system of sets satisfying the postulates:
    (1) The closed sets are Borel-sets.
    (2) The complement of a Borel-set is a Borel-set.
    (3) The sum of denumerably many Borel-sets is a Borel-set.
16. A set is *analytic* if it is the orthogonal projection of some Borel-set of a space of next higher dimension. (Every Borel-set therefore is, of course, analytic.)

Supplement to the second edition [1963]

Since the publication of the preceding paper, a number of new results have been obtained; I would like to mention those that are of special interest in connection with the foregoing discussions.

1. A. Hajnal has proved that, if $2^{\aleph_0} \neq \aleph_2$ could be derived from the axioms of set theory, so could $2^{\aleph_0} = \aleph_1$ (1956: 131). This surprising result could greatly facilitate the solution of the continuum problem, should Cantor's continuum hypothesis be demonstrable from the axioms of set theory, which, however, probably is not the case.

2. Some new consequences of, and propositions equivalent with, Cantor's hypothesis can be found in the new edition of W. Sierpiński's book (1934a, 2nd ed.). In the first edition, it had been proved that the continuum hypothesis is equivalent with the proposition that the euclidean plane is the sum of denumerably many "generalized curves" (where a generalized curve is a point set definable by an equation $y = f(x)$ in some cartesian coordinate system). In the second edition (p. 207)[24], it is pointed out that the euclidean plane can be proved to be the sum of fewer than continuum many generalized curves under the much weaker assumption that the power of the continuum is not an inaccessible number. A proof of the converse of this theorem would give some plausibility to the hypothesis $2^{\aleph_0} =$ the smallest inaccessible number $> \aleph_0$. However, great caution is called for with regard to this inference, because the paradoxical appearance in this case (like in Peano's "curves") is due (at least in part) to a transference of our geometrical intuition of curves to something which has only some of the characteristics of curves. Note that nothing of this kind is involved in the counterintuitive consequences of the continuum hypothesis mentioned on pp. [479–80].

3. C. Kuratowski has formulated a strengthening of the continuum hypothesis (1948: 131), whose consistency follows from the consistency-proof mentioned in Section 4. He then drew various consequences from this new hypothesis.

4. Very interesting new results about the axioms of infinity have been obtained in recent years (see footnotes 13 and 16).

In opposition to the viewpoint advocated in Section 4 it has been suggested (Errera 1953: 176–83) that, in case Cantor's continuum problem should turn out to be undecidable from the accepted axioms of set theory, the question of its truth would lose its meaning, exactly as the question of the truth of Euclid's fifth postulate by the proof of the consistency of non-euclidean geometry became meaningless for the mathematician. I therefore would like to point out that the situation in set theory is very different from that in geometry, both from the mathematical and from the epistemological point of view.

[24] Or Sierpiński 1951: 9. See related results in Kuratowski (1951: 15) and Sikorski 1951: 18.

In the case of the axiom of the existence of inaccessible numbers, e.g., (which can be proved to be undecidable from the von Neumann-Bernays axioms of set theory provided that it is consistent with them) there is a striking asymmetry, mathematically, between the system in which it is asserted and the one in which it is negated.[25]

Namely, the latter (but not the former) has a model which can be defined and proved to be a model in the original (unextended) system. This means that the former is an extension in a much stronger sense. A closely related fact is that the assertion (but not the negation) of the axiom implies new theorems about integers (the individual instances of which can be verified by computation). So the criterion of truth explained on p. [476] is satisfied, to some extent, for the assertion, but not for the negation. Briefly speaking, only the assertion yields a "fruitful" extension, while the negation is sterile outside its own very limited domain. Cantor's continuum hypothesis, too, can be shown to be sterile for number theory and to be true in a model constructible in the original system, whereas for some other assumption about the power of the continuum this perhaps is not so. On the other hand neither one of those asymmetries applies to Euclid's fifth postulate. To be more precise, both it and its negation are extensions in the weak sense.

As far as the epistemological situation is concerned, it is to be said that by a proof of undecidability a question loses its meaning only if the system of axioms under consideration is interpreted as a hypothetico-deductive system; i.e., if the meanings of the primitive terms are left undetermined. In geometry, e.g., the question as to whether Euclid's fifth postulate is true retains its meaning if the primitive terms are taken in a definite sense, i.e., as referring to the behavior of rigid bodies, rays of light, etc. The situation in set theory is similar, the difference is only that, in geometry, the meaning usually adopted today refers to physics rather than to mathematical intuition and that, therefore, a decision falls outside the range of mathematics. On the other hand, the objects of transfinite set theory, conceived in the manner explained on pp. [474–5] and in footnote 11, clearly do not belong to the physical world and even their indirect connection with physical experience is very loose (owing primarily to the fact that set-theoretical concepts play only a minor role in the physical theories of today).

But, despite their remoteness from sense experience, we do have some-

[25] The same asymmetry also occurs on the lowest levels of set theory, where the consistency of the axioms in question is less subject to being doubted by skeptics.

thing like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them and, moreover, to believe that a question not decidable now has meaning and may be decided in the future. The set-theoretical paradoxes are hardly any more troublesome for mathematics than deceptions of the senses are for physics. That new mathematical intuitions leading to a decision of such problems as Cantor's continuum hypothesis are perfectly possible was pointed out earlier (pp. [476–7]).

It should be noted that mathematical intuition need not be conceived of as a faculty giving an *immediate* knowledge of the objects concerned. Rather it seems that, as in the case of physical experience, we *form* our ideas also of those objects on the basis of something else which *is* immediately given. Only this something else here is *not*, or not primarily, the sensations. That something besides the sensations actually is immediately given follows (independently of mathematics) from the fact that even our ideas referring to physical objects contain constituents qualitatively different from sensations or mere combinations of sensations, e.g., the idea of object itself, whereas, on the other hand, by our thinking we cannot create any qualitatively new elements, but only reproduce and combine those that are given. Evidently the "given" underlying mathematics is closely related to the abstract elements contained in our empirical ideas.[26] It by no means follows, however, that the data of this second kind, 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 in us may be due to another kind of relationship between ourselves and reality.

However, the question of the objective existence of the objects of mathematical intuition (which, incidentally, is an exact replica of the question of the objective existence of the outer world) is not decisive for the problem under discussion here. The mere psychological fact of the existence of an intuition which is sufficiently clear to produce the axioms of set theory and an open series of extensions of them suffices to give meaning to the question of the truth or falsity of propositions like

[26] Note that there is a close relationship between the concept of set explained in footnote 14 and the categories of pure understanding in Kant's sense. Namely, the function of both is "synthesis," i.e., the generating of unities out of manifolds (e.g., in Kant, of the idea of *one* object out of its various aspects).

Cantor's continuum hypothesis. What, however, perhaps more than anything else, justifies the acceptance of this criterion of truth in set theory is the fact that continued appeals to mathematical intuition are necessary not only for obtaining unambiguous answers to the questions of transfinite set theory, but also for the solution of the problems of finitary number theory[27] (of the type of Goldbach's conjecture),[28] where the meaningfulness and unambiguity of the concepts entering into them can hardly be doubted. This follows from the fact that for every axiomatic system there are infinitely many undecidable propositions of this type.

It was pointed out earlier (p. [477]) 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. This criterion, however, though it may become decisive in the future, cannot yet be applied to the specifically set-theoretical axioms (such as those referring to great cardinal numbers), because very little is known about their consequences in other fields. The simplest case of an application of the criterion under discussion arises when some set-theoretical axiom has number-theoretical consequences verifiable by computation up to any given integer. On the basis of what is known today, however, it is not possible to make the truth of any set-theoretical axiom reasonably probable in this manner.

Postscript

Shortly after the completion of the manuscript of this paper the question of whether Cantor's Continuum Hypothesis is provable from the von Neumann-Bernays axioms of set theory (the axiom of choice included) was settled in the negative by Paul J. Cohen (1963a, 1964). It turns out that for a wide range of $\aleph_\tau$, the equality $2^{\aleph_0} = \aleph_\tau$ is consistent and an extension in the weak sense (that is, it implies no new number-theoretical theorems). Whether for a suitable concept of "standard" definition there exist definable $\aleph_\tau$ not excluded by König's theorem (see p. [472] above) for which this is not so is still an open question (of course, it must be assumed that the existence of the $\aleph_\tau$ in question is either demonstrable or has been postulated).

[27] Unless one is satisfied with inductive (probable) decisions, such as verifying the theorem up to very great numbers, or more indirect inductive procedures (see pp. [478, 485]).

[28] I.e., universal propositions about integers which can be decided in each individual instance.

The iterative conception of set

GEORGE BOOLOS

A set, according to Cantor, is "any collection ... into a whole of definite, well-distinguished objects ... of our intuition or thought."[1] Cantor also defined a set as a "many, which can be thought of as one, i.e., a totality of definite elements that can be combined into a whole by a law."[2] One might object to the first definition on the grounds that it uses the concepts of *collection* and *whole,* which are notions no better understood than that of *set,* that there ought to be sets of objects that are not objects of our thought, that 'intuition' is a term laden with a theory of knowledge that no one should believe, that *any* object is "definite," that there should be sets of ill-distinguished objects, such as waves and trains, etc., etc. And one might object to the second on the grounds that 'a many' is ungrammatical, that if something is "a many" it should hardly be thought of as one, that *totality* is as obscure as *set,* that it is far from clear how laws can combine anything into a whole, that there ought to be other combinations into a whole than those effected by "laws," etc., etc. But it cannot be denied that Cantor's definitions could be used by a person to identify and gain some understanding of the sort of object of which Cantor wished to treat. Moreover, they do suggest – although, it must be conceded, only very faintly – two important characteristics of sets: that a set is "determined" by its elements in the sense that sets with exactly the same elements are identical, and that, in a sense, the clarification of which is one of the principal objects of the theory whose rationale we shall give, the elements of a set are "prior to" it.

It is not to be presumed that the concepts of *set* and *member of* can be explained or defined by means of notions that are simpler or conceptually more basic. However, as a theory about sets might itself provide the sort of elucidation about sets and membership that good definitions

Reprinted with the kind permission of the author and the editors from the *Journal of Philosophy* 68 (1971): 215–32.

[1] "Unter einer 'Menge' verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die 'Elemente' von M genannt werden) zu einem Ganzen" (Cantor 1932: 282).

[2] "... jedes Viele, welches sich als Eines denken lässt, d.h. jeden Inbegriff bestimmter Elemente, welcher durch ein Gesetz zu einem Ganzen verbunden werden kann" (Cantor 1932: 204).

might be hoped to offer, there is no reason for such a theory to begin with, or even contain, a definition of 'set'. That we are unable to give informative definitions of *not* or *for some* does not and should not prevent the development of quantificational logic, which provides us with significant information about these concepts.

I. Naive set theory

Here is an idea about sets that might occur to us quite naturally, and is perhaps suggested by Cantor's definition of a set as a totality of definite elements that can be combined into a whole *by a law.*

By the law of excluded middle, any (one-place) predicate in any language either applies to a given object or does not. So, it would seem, to any predicate there correspond two sorts of thing: the sort of thing to which the predicate applies (of which it is true) and the sort of thing to which it does not apply. So, it would seem, for any predicate there is a set of all and only those things to which it applies (as well as a set of just those things to which it does not apply). Any set whose members are exactly the things to which the predicate applies – by the axiom of extensionality, there cannot be *two* such sets – is called the *extension* of the predicate. Our thought might therefore be put: "Any predicate has an extension." We shall call this proposition, together with the argument for it, the *naive conception of set.*

The argument has great force. How could there *not* be a collection, or set, of just those things to which any given predicate applied? Isn't anything to which a predicate applies similar to all other things to which it applies in precisely the respect that it applies to them; and how could there fail to be a set of all things similar to one another in this respect? Wouldn't it be extremely implausible to say, of any particular predicate one might consider, that there weren't two kinds of thing it determined, namely, a kind of thing of which it is true, and a kind of thing of which it is not true? And why should one not take these kinds of things to be sets? Aren't kinds sets? If not, what is the difference?

Let us denote by '$\mathcal{K}$' a certain standardly formalized first-order language, whose variables range over all sets and individuals (= non-sets), and whose nonlogical constants are a one-place predicate letter 'S' abbreviating 'is a set', and a two-place predicate letter '$\in$', abbreviating 'is a member of'. Which sentences of this language, together with their consequences, do we believe state truths about sets? Otherwise put, which formulas of $\mathcal{K}$ should we take as axioms of a set theory on the strength of our beliefs about sets?

If the naive conception of set is correct, there should (at least) be a set of just those things to which ϕ applies, if ϕ is a formula of $\mathcal{K}$. So (the universal closure of) $\ulcorner(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow \phi))\urcorner$ should express a truth about sets (if no occurrence of 'y' in ϕ is free).

We call the theory whose axioms are the axiom of extensionality (to which we later recur), i.e., the sentence

$$(x)(y)(Sx \,\&\, Sy \,\&\, (z)(z \in x \leftrightarrow z \in y) \rightarrow x = y)$$

and all formulas $\ulcorner(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow \phi))\urcorner$ (where 'y' does not occur free in ϕ) *naive set theory.*

Some of the axioms of naive set theory are the formulas

$$(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow x \neq x))$$
$$(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow (x = z \vee x = w)))$$
$$(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow (\exists w)(x \in w \,\&\, w \in z)))$$
$$(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow (Sx \,\&\, x = x)))$$

The first of these formulas states that there is a set that contains no members. By the axiom of extensionality, there can be at most one such set. The second states that there is a set whose sole members are z and w; the third, that there is a set whose members are just the members of members of z.

The last, which states that there is a set that contains all sets whatsoever, is rather anomalous; for if there is a set that contains all sets, a universal set, that set contains itself, and perhaps the mind ought to boggle at the idea of something's *containing* itself. Nevertheless, naive set theory is simple to state, elegant, initially quite credible, and natural in that it articulates a view about sets that might occur to one quite naturally. Alas, it is inconsistent.

Proof of the inconsistency of naive set theory
(Russell's paradox)

No set can contain all and only those sets which do not contain themselves. For if any such set existed, if it contained itself, then, as it contains *only* those sets which do not contain themselves, it would not contain itself; but if it did not contain itself, then, as it contains *all* those sets which do not contain themselves, it would contain itself. Thus any such set would have to contain itself if and only if it did not contain itself. Consequently, there is no set that contains all and only those sets which do not contain themselves.

This argument, which uses no axioms of naive set theory, or any other set theory, shows that the sentence

$$\sim(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow (Sx \,\&\, \sim x \in x)))$$

is *logically valid* and, hence, is a theorem of any theory that is expressed in $\mathcal{K}$. But one of the axioms and, hence, one of the theorems, of naive set theory is the sentence

$$(\exists y)(Sy \,\&\, (x)(x \in y \leftrightarrow (Sx \,\&\, \sim x \in x)))$$

Naive set theory is therefore inconsistent.

II. The iterative conception of set

Faced with the inconsistency of naive set theory, one might come to believe that any decision to adopt a system of axioms about sets would be *arbitrary* in that no explanation could be given why the particular system adopted had any greater claim to describe what we conceive sets and the membership relation to be like than some other system, perhaps incompatible with the one chosen. One might think that no answer could be given to the question: why adopt *this* particular system rather than that or this other one? One might suppose that any apparently consistent theory of sets would have to be unnatural in some way or fragmentary, and that, if consistent, its consistency would be due to certain provisions that were laid down for the express purpose of avoiding the paradoxes that show naive set theory inconsistent, but that lack any independent motivation.

One might imagine all this; but there is another view of sets: the *iterative conception of set,* as it is sometimes called, which often strikes people as entirely natural, free from artificiality, not at all ad hoc, and one they might perhaps have formulated themselves.

It is, perhaps, no more natural a conception than the naive conception, and certainly not quite so simple to describe. On the other hand, it is, as far as we know, consistent: not only are the sets whose existence would lead to contradiction not assumed to exist in the axioms of the theories that express the iterative conception, but the more than fifty years of experience that practicing set theorists have had with this conception have yielded a good understanding of what can and what cannot be proved in these theories, and at present there just is no suspicion at all that they are inconsistent.[3]

[3]The conception is well known among logicians; a rather different version of it is sketched in Shoenfield (1967: chap. 10). I learned of it principally from Putnam, Kripke, and Donald Martin. Authors of set-theory texts either omit it or relegate it to back pages; philosophers, in the main, seem to be unaware of it, or of the preeminence of ZF, which may be said to embody it. It is due primarily to Zermelo and Russell.

The standard, first-order theory that expresses the iterative conception of set as fully as a first-order theory in the language $\mathfrak{L}$ of set theory[4] can, is known as *Zermelo-Fraenkel set theory,* or 'ZF' for short. There are other theories besides ZF that embody the iterative conception: one of them, Zermelo set theory, or "Z", which will occupy us shortly, is a *subsystem* of ZF in the sense that any theorem of Z is also a theorem of ZF; two others, von-Neumann-Bernays-Gödel set theory and Morse-Kelley set theory, are supersystems (or extensions) of ZF, but they are most commonly formulated as second-order theories.

Other theories of sets, incompatible with ZF, have been proposed.[5] These theories appear to lack a motivation that is independent of the paradoxes in the following sense: they are not, as Russell has written, "such as even the cleverest logician would have thought of if he had not known of the contradictions" (1959: 80). A final and satisfying resolution to the set-theoretical paradoxes cannot be embodied in a theory that blocks their derivation by artificial technical restrictions on the set of axioms that are imposed *only because* paradox would otherwise ensue; these other theories survive only through such artificial devices. ZF alone (together with its extensions and subsystems) is not only a consistent (apparently) but also an independently motivated theory of sets: there is, so to speak, a "thought behind it" about the nature of sets which might have been put forth even if, impossibly, naive set theory had been consistent. The thought, moreover, can be described in a rough, but informative way without first stating the theory the thought is behind.

In order to see why a conception of set other than the naive conception might be desired even if the naive conception were consistent, let us take another look at naive set theory and the anomalousness of its axiom, '$(\exists y)(Sy \; \& \; (x)(x \in y \leftrightarrow (Sx \; \& \; x=x)))$'.

According to this axiom there is a set that contains all sets, and therefore there is a set that contains itself. It is important to realize how odd the idea of something's containing itself is. Of course a set can and must *include* itself (as a subset). But *contain* itself? Whatever tenuous hold on the concepts of *set* and *member* were given one by Cantor's definitions of 'set' and one's ordinary understanding of 'element', 'set', 'collection', etc. is altogether lost if one is to suppose that some sets are members of themselves. The idea is paradoxical not in the sense that it is contradictory to suppose that some set is a member of itself, for, after all, '$(\exists x)(Sx \; \& \; x \in x)$' is obviously consistent, but that if one understands '$\in$' as meaning 'is a member of', it is very, very peculiar to suppose it

[4] $\mathfrak{L}$ contains (countably many) variables, ranging over (pure) sets, '=', and '$\in$', which is its sole nonlogical constant.

[5] For example, Quine's systems NF and ML.

true. For when one is told that a set is a collection into a whole of definite elements of our thought, one thinks: Here are some things. Now we bind them up into a whole.[6] *Now* we have a set. We don't suppose that what we come up with after combining some elements into a whole could have been one of the very things we combined (not, at least, if we are combining two or more elements).

If $(\exists x)(Sx \; \& \; x \in x)$, then $(\exists x)(\exists y)(Sx \; \& \; Sy \; \& \; x \in y \; \& \; y \in x)$. The supposition that there are sets x and y each of which belongs to the other is almost as strange as the supposition that some set is a self-member. There is of course an infinite sequence of such cyclical pathologies: $(\exists x)(\exists y)(\exists z)(Sx \; \& \; Sy \; \& \; Sz \; \& \; x \in y \; \& \; y \in z \; \& \; z \in x)$, etc. Only slightly less pathological are the suppositions that there is an ungrounded set,[7] or that there is an infinite sequence of sets $x_0, x_1, x_2, \ldots$, each term of which belongs to the previous one.

There does not seem to be any argument that is guaranteed to persuade someone who really does not see the peculiarity of a set's belonging to itself, or to one of its members, etc., that these states of affairs are peculiar. But it is in part the sense of their oddity that has led set-theorists to favor conceptions of set, such as the iterative conception, according to which what they find odd does not occur.

We describe this conception now. Our description will have three parts. The first is a rough statement of the idea. It contains such expressions as 'stage', 'is formed at', 'earlier than', 'keep on going', which must be exorcised from any formal theory of sets. From the rough description it sounds as if sets were continually being created, which is not the case. In the second part, we present an axiomatic theory which partially formalizes the idea roughly stated in the first part. For reference, let us call this theory the *stage theory.* The third part consists in a derivation from the stage theory of the axioms of a theory of sets. These axioms are formulas of $\mathfrak{L}$, the language of set theory, and contain none of the metaphorical expressions which are employed in the rough statement and of which abbreviations are found in the language in which the stage theory is expressed.

Here is the idea, roughly stated:

A set is any collection that is formed at some stage of the following process: Begin with individuals (if there are any). An individual is an object that is not a set; individuals do not contain members. At stage zero (we count from zero instead of one) form all possible collections of

[6] We put a "lasso" around them, in a figure of Kripke's.

[7] x is *ungrounded* if x belongs to some set z, each of whose members has a member in common with z.

individuals. If there are no individuals, only one collection, the null set, which contains no members, is formed at this 0th stage. If there is only one individual, two sets are formed: the null set and the set containing just that one individual. If there are two individuals, four sets are formed; and in general, if there are n individuals, 2^n sets are formed. Perhaps there are infinitely many individuals. Still, we assume that one of the collections formed at stage zero is the collection of all individuals, however many of them there may be.

At stage one, form all possible collections of individuals and sets formed at stage zero. If there are any individuals, at stage one some sets are formed that contain both individuals and sets formed at stage zero. Of course some sets are formed that contain only sets formed at stage zero. At stage two, form all possible collections of individuals, sets formed at stage zero, and sets formed at stage one. At stage three, form all possible collections of individuals and sets formed at stages zero, one, and two. At stage four, form all possible collections of individuals and sets formed at stages zero, one, two, and three. Keep on going in this way, at each stage forming all possible collections of individuals and sets formed at earlier stages.

Immediately after all of stages zero, one, two, three,..., there is a stage; call it stage omega. At stage omega, form all possible collections of individuals formed at stages zero, one, two,.... One of these collections will be the set of *all* sets formed at stages zero, one, two,....

After stage omega there is a stage omega plus one. At stage omega plus one form all possible collections of individuals and sets formed at stages zero, one, two..., and omega. At stage omega plus two form all possible collections of individuals and sets formed at stages zero, one, two,..., omega, and omega plus one. At stage omega plus three form all possible collections of individuals and sets formed at earlier stages. Keep on going in this way.

Immediately after all of stages zero, one, two,..., omega, omega plus one, omega plus two,..., there is a stage, call it stage omega plus omega (or omega times two). At stage omega plus omega form all possible collections of individuals and sets formed at earlier stages. At stage omega plus omega plus one.... ...

...omega plus omega plus omega (or omega times three)...

...(omega times four)...

...omega times omega... ...

Keep on going in this way....

According to this description, sets are formed over and over again: in fact, according to it, a set is formed at every stage later than that at which it is first formed. We could continue to say this if we liked; instead we

shall say that a set is formed only once, namely, at the earliest stage at which, on our old way of speaking, it would have been said to be formed.

That is a rough statement of the iterative conception of set. According to this conception, no set belongs to itself, and hence there is no set of all sets; for every set is formed at some earliest stage, and has as members only individuals or sets formed at still earlier stages. Moreover, there are not two sets x and y, each of which belongs to the other. For if y belonged to x, y would have had to be formed at an earlier stage than the earliest stage at which x was formed, and if x belonged to y, x would have had to be formed at an earlier stage than the earliest stage at which y was formed. So x would have had to be formed at an earlier stage than the earliest stage at which it was formed, which is impossible. Similarly, there are no sets x, y, and z such that x belongs to y, y to z, and z to x. And in general, there are no sets $x_0, x_1, x_2, \ldots, x_n$ such that x_0 belongs to x_1, x_1 to $x_2, \ldots, x_{n-1}$ to x_n, and x_n to x_0. Furthermore it would appear that there is no sequence of sets $x_0, x_1, x_2, x_3, \ldots$ such that x_1 belongs to x_0, x_2 belongs to x_1, x_3 belongs to x_2, and so forth. Thus, if sets are as the iterative conception has them, the anomalous situations do not arise in which sets belong to themselves or to others that in turn belong to them.

The sets of which ZF in its usual formulation speaks ("quantifies over") are not all the sets there are, if we assume that there are some individuals, but only those which are formed at some stage under the assumption that there are no individuals. These sets are called *pure* sets. All members of a pure set are pure sets, and any set, all of whose members are pure, is itself pure. It may not be obvious that any pure sets are ever formed, but the set Λ, which contains no members at all, is pure, and is formed at stage 0. $\{\Lambda\}$ and $\{\{\Lambda\}\}$ are also both pure and are formed at stages 1 and 2, respectively. There are many others. From now on, we shall use the word 'set' to mean 'pure set'.

Let us now try to state a theory, the stage theory, that precisely expresses much, but not all, of the content of the iterative conception. We shall use a language, $\mathcal{J}$, in which there are two sorts of variables: variables 'x', 'y', 'z', 'w',..., which range over sets, and variables 'r', 's', 't', which range over stages. In addition to the predicate letters '$\in$' and '$=$' of $\mathcal{L}$, $\mathcal{J}$ also contains two new two-place predicate letters 'E', read 'is earlier than', and 'F', read 'is formed at'. The rules of formation of $\mathcal{J}$ are perfectly standard.

Here are some axioms governing the sequence of stages:

(I) $(x){\sim}s\mathrm{E}s$ (No stage is earlier than itself.)

(II) $(r)(s)(t)((r\mathrm{E}s \;\&\; s\mathrm{E}t) \rightarrow r\mathrm{E}t)$ (*Earlier than* is transitive.)

(III) $(s)(t)(s\mathrm{E}t \vee s=t \vee t\mathrm{E}s)$ (*Earlier than* is connected.)
(IV) $(\exists s)(t)(t \neq s \rightarrow s\mathrm{E}t)$ (There is an earliest stage.)
(V) $(s)(\exists t)(s\mathrm{E}t \mathbin{\&} (r)(r\mathrm{E}t \rightarrow (r\mathrm{E}s \vee r=s)))$ (*Immediately* after any stage there is another.)

Here are some axioms describing when sets and their members are formed:

(VI) $(\exists s)((\exists t)t\mathrm{E}s \mathbin{\&} (t)(t\mathrm{E}s \rightarrow (\exists r)(t\mathrm{E}r \mathbin{\&} r\mathrm{E}s)))$ (There is a stage, not the earliest one, which is not immediately after any one stage. In the rough description, stage omega was such a stage.)
(VII) $(x)(\exists s)(x\mathrm{F}s \mathbin{\&} (t)(x\mathrm{F}t \rightarrow t=s))$ (Every set is formed at some unique stage.)
(VIII) $(x)(y)(s)(t)((y \in x \mathbin{\&} x\mathrm{F}s \mathbin{\&} y\mathrm{F}t) \rightarrow t\mathrm{E}s)$ (Every member of a set is formed *before,* i.e., at an earlier stage than, the set.)
(IX) $(x)(s)(t)(x\mathrm{F}s \mathbin{\&} t\mathrm{E}s \rightarrow (\exists y)(\exists r)(y \in x \mathbin{\&} y\mathrm{F}r \mathbin{\&} (t=r \vee t\mathrm{E}r)))$ (If a set is formed at a stage, then, at or after any earlier stage, at least one of its members is formed. So it never happens that all members of a set are formed before some stage, but the set is not formed at that stage, if it has not been formed already.)

We may capture part of the content of the idea that at any stage every *possible* collection (or set) of sets formed at earlier stages is formed (if it has not yet been formed) by taking as axioms all formulas $\ulcorner(s)(\exists y)(x)(x \in y \leftrightarrow (\chi \mathbin{\&} (\exists t)(t\mathrm{E}s \mathbin{\&} x\mathrm{F}t)))\urcorner$, where χ is a formula of the language $\mathcal{J}$ in which no occurrence of 'y' is free. Any such axiom will say that for any stage there is a set of just those sets to which χ applies that are formed before that stage. Let us call these axioms *specification axioms.*

There is still one important feature contained in our rough description that has not yet been expressed in the stage theory: the analogy between the way sets are *inductively generated* by the procedure described in the rough statement and the way the natural numbers $0, 1, 2, \ldots$ are inductively generated from 0 by the repeated application of the successor operation. One way to characterize this feature is to assert a suitable induction principle concerning sets and stages; for, as Frege, Dedekind, Peano, and others have enabled us to see, the content of the idea that objects of a certain kind are inductively generated in a certain way is just the proposition than an appropriate induction principle holds of those objects.

The principle of mathematical induction, the induction principle governing the natural numbers, has two forms, which are interderivable on

certain assumptions about the natural numbers. The first version of the principle is the statement

$$(P)[(P0 \mathbin{\&} (n)[Pn \rightarrow PSn]) \rightarrow (n)Pn]$$

which may be read, 'If 0 has a property and if whenever a natural number has the property its successor does, then every natural number has the property'. The second version is the statement

$$(P)[(n)((m)[m<n \rightarrow Pm] \rightarrow Pn) \rightarrow (n)Pn]$$

It may be read, '*If each natural number has a property provided that all smaller natural numbers have it, then every natural number has the property*'.

The induction principle about sets and stages that we should like to assert is modeled after the second form of the principle of mathematical induction. Let us say that a stage *s is covered by* a predicate if the predicate applies to every set formed at *s*. Our analogue for sets and stages of the second form of mathematical induction says that *if each stage is covered by a predicate provided that all earlier stages are covered by it, then every stage is covered by the predicate.* The full force of this assertion can be expressed only with a second-order quantifier. However, we can capture some of its content by taking as axioms all formulas

$$\ulcorner(s)((t)(t\mathrm{E}s \rightarrow (x)(x\mathrm{F}t \rightarrow \theta)) \rightarrow (x)(x\mathrm{F}s \rightarrow \chi)) \rightarrow (s)(x)(x\mathrm{F}s \rightarrow \chi)\urcorner$$

where χ is a formula of $\mathcal{J}$ that contains no occurrences of 't' and θ is just like χ except for containing a free occurrence of 't' wherever χ contains a free occurrence of 's'. [Observe that '$(x)(x\mathrm{F}s \rightarrow \chi)$' says that χ applies to every set formed at stage s and, hence, that s is covered by χ.] We call these axioms *induction axioms.*

III. Zermelo set theory

We complete the description of the iterative conception of set by showing how to derive the axioms of a theory of sets from the stage theory. The axioms we derive speak only about sets and membership: they are formulas of $\mathcal{L}$.

The axiom of the null set: $(\exists y)(x){\sim}x \in y$. (There is a set with no members.)

Derivation. Let $\chi =$ '$x=x$'. Then

$$(s)(\exists y)(x)(x \in y \leftrightarrow (x=x \mathbin{\&} (\exists t)(t\mathrm{E}s \mathbin{\&} x\mathrm{F}t)))$$

is a specification axiom, according to which, for any stage, there is a set of all sets formed at earlier stages. As there is an earliest stage, stage 0, before which no sets are formed, there is a set that contains no members. Note that, by axiom (IX) of the stage theory, any set with no members is formed at stage 0; for if it were formed later, it would have to have a member (that was formed at or after stage 0).

The axiom of pairs: $(z)(w)(\exists y)(x)(x \in y \leftrightarrow (x=z \lor x=w))$. (For any sets z and w, not necessarily distinct, there is a set whose sole members are z and w.)

Derivation. Let $\chi=$'$(x=z \lor x=w)$'. Then

$$(s)(\exists y)(x)(x \in y \leftrightarrow ((x=z \lor x=w) \;\&\; (\exists t)(t\mathrm{E}s \;\&\; x\mathrm{F}t)))$$

is a specification axiom, according to which, for any stage, there is a set of all sets formed at earlier stages that are identical with either z or w. Any set is formed at some stage. Let r be the stage at which z is formed; s, the stage at which w is formed. Let t be a stage later than both r and s. Then there is a set of all sets formed at stages earlier than t that are identical with z or w. So there is a set containing just z and w.

The axiom of unions: $(z)(\exists y)(x)(x \in y \leftrightarrow (\exists w)(x \in w \;\&\; w \in z))$. (For any set z, there is a set whose members are just the members of members of z.)

Derivation. '$(s)(\exists y)(x)(x \in y \leftrightarrow ((\exists w)(x \in w \& w \in z) \& (\exists t)(t\mathrm{E}s \& x\mathrm{F}t)))$' is a specification axiom, according to which, for any stage, there is a set of all members of members of z formed at earlier stages. Let s be the stage at which z is formed. Every member of z is formed before s, and hence every member of a member of z is also formed before s. Thus there is a set of all members of members of z.

The power-set axiom: $(z)(\exists y)(x)(x \in y \leftrightarrow (w)(w \in x \rightarrow w \in z))$. (For any set z, there is a set whose members are just the subsets of z.)

Derivation. '$(s)(\exists y)(x)(x \in y \leftrightarrow ((w)(w \in x \rightarrow w \in z) \& (\exists t)(t\mathrm{E}s \& x\mathrm{F}t)))$' is a specification axiom, according to which, for any stage, there is a set of all subsets of z formed at earlier stages. Let t be the stage at which z is formed and let s be the next later stage. If x is a subset of z, then x is formed before s. For otherwise, by axiom (IX), there would be a member of x that was formed at or after t and, hence, that was not a member of z.

So there is a set of all subsets of z formed before s, and hence a set of all subsets of z.

The axiom of infinity:

$$(\exists y)((\exists x)(x \in y \;\&\; (z) \sim z \in x) \;\&\; (x)(x \in y \rightarrow (\exists z)(z \in y \;\&\; (w)(w \in z \leftrightarrow (w \in x \lor w=x)))))$$

(Call a set null if it has no members. Call z a successor of x if the members of z are just those of x and x itself. Then there is a set which contains a null set and which contains a successor of any set it contains.)

Derivation. Let us first observe that every set x has a successor. For let y be a set containing just x and x (axiom of pairs), and let w be a set containing just x and y (axiom of pairs again), and let z contain just the members of members of w (axiom of unions). Then z is a successor of x, for its members are just x and x's members. Next, note that if z is a successor of x, x is formed at r, and t is the next stage after r, then z is formed at t. For every member of z is formed before t. So z is formed at or before t, by axiom (IX). But x, which is in z, is formed at r. So z cannot be formed at or before r. So z cannot be formed before t. Now, by axiom (VI), there is a stage s, not the earliest one, which is not immediately after any stage. '$(s)(\exists y)(x)(x \in y \leftrightarrow (x=x \;\&\; (\exists t)(t\mathrm{E}s \;\&\; x\mathrm{F}t)))$' is a specification axiom, according to which, for any stage, there is a set of all sets formed at earlier stages. So there is a set y of all sets formed before s. y thus contains all sets formed at stage 0, and hence contains a null set. And if y contains x, y contains all successors of x (and there are some), for all these are formed at stages immediately after stages before s and, hence, at stages themselves before s.

Axioms of separation (Aussonderungsaxioms): All formulas

$$\ulcorner(z)(\exists y)(x)(x \in y \leftrightarrow (x \in z \;\&\; \phi))\urcorner$$

where ϕ is a formula of L in which no occurrence of 'y' is free.

Derivation. If ϕ is a formula of $\mathcal{L}$ in which no occurrence of 'y' is free, then $\ulcorner(s)(\exists y)(x)(x \in y \leftrightarrow ((x \in z \;\&\; \phi) \;\&\; (\exists t)(t\mathrm{E}s \;\&\; x\mathrm{F}t)))\urcorner$ is a specification axiom, which we may read, 'for any stage s, there is a set of all sets formed at earlier stages, which belong to z and to which ϕ applies. Let s be the stage at which z is formed. All members of z are formed before s. So, for any z, there is a set of just those members of z to which ϕ applies, which we would write, $\ulcorner(z)(\exists y)(x)(x \in y \leftrightarrow (x \in z \;\&\; \phi))\urcorner$. A formal derivation of an Aussonderungsaxiom would use the specification axiom described and axioms (VII) and (VIII) of the stage theory.

Axioms of regularity: All formulas

$$\ulcorner(\exists x)\phi \to (\exists x)(\phi \;\&\; (y)(y\in x \to {\sim}\psi))\urcorner,$$

where ϕ does not contain 'y' at all and ψ is just like ϕ except for containing an occurrence of 'y' wherever ϕ contains a free occurrence of 'x'.

Derivation. The idea: Suppose ϕ applies to some set x'. x' is formed at some stage. That stage is therefore not covered by $\ulcorner{\sim}\phi\urcorner$. By an induction axiom, there is then a stage s not covered by $\ulcorner{\sim}\phi\urcorner$, although all stages earlier than s are covered by $\ulcorner{\sim}\phi\urcorner$. Since s is not covered by $\ulcorner{\sim}\phi\urcorner$, there is an x, formed at s, to which $\ulcorner{\sim}\phi\urcorner$ does not apply, i.e., to which ϕ applies. If y is in x, however, y is formed before s, and hence the stage at which it is formed is covered by $\ulcorner{\sim}\phi\urcorner$. So $\ulcorner{\sim}\phi\urcorner$ applies to y (which is what $\ulcorner{\sim}\psi\urcorner$ says).

For a formal derivation, contrapose, reletter, and simplify the induction axiom

$$\ulcorner(s)((t)(t\mathrm{E}s \to (x)(x\mathrm{F}t \to {\sim}\phi)) \to (x)(x\mathrm{F}s \to {\sim}\phi)) \\ \to (s)(x)(x\mathrm{F}s \to {\sim}\phi)\urcorner$$

so as to obtain

$$\ulcorner(\exists s)(\exists x)(x\mathrm{F}s \;\&\; \phi) \to (\exists s)(\exists x)(x\mathrm{F}s \;\&\; \phi \;\&\; (y)(t)(t\mathrm{E}s \;\&\; y\mathrm{F}t \to {\sim}\psi))\urcorner$$

Assume $\ulcorner(\exists x)\phi\urcorner$. Use axiom (VII) and modus ponens to obtain

$$\ulcorner(\exists s)(\exists x)(x\mathrm{F}s \;\&\; \phi \;\&\; (y)(t)(t\mathrm{E}s \;\&\; y\mathrm{F}t \to {\sim}\psi))\urcorner$$

Use axioms (VII) and (VIII) to obtain $\ulcorner(\exists x)(\phi \;\&\; (y)(y\in x \to {\sim}\psi))\urcorner$ from this.

The axioms of regularity (partially) express the analogue for sets of the version of mathematical induction called the *least-number principle:* if there is a number that has a property, then there is a least number with that property. The analogue itself has been called the *principle of set-theoretical induction.*[8] Here is an application of set-theoretical induction.

Theorem: No set belongs to itself.

Proof. Suppose that some set belongs to itself, i.e., that $(\exists x)x\in x$.

$$(\exists x)x\in x \to (\exists x)(x\in x \;\&\; (y)(y\in x \to {\sim}y\in y))$$

[8] By Tarski, among others.

is an axiom of regularity. By modus ponens, then, some set x belongs to itself though no member of x (not even x) belongs to itself. This is a contradiction.

The axioms whose derivations we have given are those statements which are often taken as axioms of ZF and which are deducible from all (sufficiently strong[9]) theories that can fairly be called formalizations of the iterative conception, as roughly described. (The axiom of extensionality has a special status, which we discuss below.) Other axioms than those we have given could have been taken as axioms of the stage theory. For example, we could have fairly taken as an axiom a statement asserting the existence of a stage, not immediately later than any stage, but later than some stage that is itself neither the earliest stage nor immediately later than any stage. Such an axiom would have enabled us to deduce a stronger axiom of infinity than the one whose derivation we have given, but this stronger statement is not commonly taken as an axiom of ZF. We could also have derived other statements from the stage theory, such as the statement that no set belongs to any of its members, but this statement is never taken as an axiom of ZF. We do not believe that the axioms of replacement or choice can be inferred from the iterative conception.

One of the axioms of regularity,

$$(z)((\exists x)x\in z \to (\exists x)(x\in z \;\&\; (y)(y\in x \to {\sim}y\in z)))$$

is sometimes called *the* axiom of regularity; in the presence of other axioms of ZF, all the other axioms of regularity follow from it. The name 'Zermelo set theory' is perhaps most commonly given to the theory whose axioms are '$(x)(y)((z)(z\in x \leftrightarrow x\in y) \to x=y)$', i.e., the axiom of extensionality, and the axioms of the null set, pairs, and unions, the power-set axiom, the axiom of infinity, all the Aussonderungsaxioms, and the axiom of regularity.[10] With the exception of the axiom of extensionality, then, all the axioms of Zermelo set theory follow from the stage theory.

IV. Zermelo-Fraenkel set theory

The axioms of replacement. ZF is the theory whose axioms are those of Zermelo set theory and all axioms of replacement.[11] A formula of $\mathcal{L}$ is an

[9] 'Sufficiently strong' may here be taken to mean "at least as strong as the stage theory."

[10] Zermelo (1908) took as axioms versions of the axioms of extensionality, the null set, pairs (and unit set), unions, the power-set axiom, the axiom of infinity, the Aussonderungs-axioms, and the axiom of choice.

[11] Sometimes the axiom of choice is also considered one of the axioms of ZF.

axiom of replacement if it is the translation into $\mathcal{L}$ of the result "substituting" a formula of $\mathcal{L}$ for 'F' in

$$F \text{ is a function} \rightarrow (z)(\exists y)(x)(x \in y \leftrightarrow (\exists w)(w \in z \;\&\; F(w) = x))$$

There is an extension of the stage theory from which the axioms of replacement could have been derived. We could have taken as axioms all instances (that can be expressed in $\mathcal{J}$) of a principle which may be put, 'If each set is correlated with at least one stage (no matter how), then for any set z there is a stage s such that for each member w of z, s is later than some stage with which w is correlated'. This *bounding* or *cofinality* principle is an attractive further thought about the interrelation of sets and stages, but it does seem to us to be a *further* thought, and not one that can be said to have been meant in the rough description of the iterative conception. For that there are exactly ω_1 stages does not seem to be excluded by anything said in the rough description; it would seem that $R\omega_1$ (see below) is a model for any statement of $\mathcal{L}$ that can (fairly) be said to have been implied by the rough description, and not all of the axioms of replacement hold in $R\omega_1$.[12] Thus the axioms of replacement do not seem to us to follow from the iterative conception.

Adding the axioms of replacement to those of Zermelo set theory enables us to define a sequence of sets, $\{R_\alpha\}$, with which the stages of the stage theory may be identified. Suppose we put $R_0 =$ the null set; $R_{\alpha+1} = R_\alpha \cup$ the power set of R_α, and $R_\lambda = \bigcup_{\beta<\lambda} R_\beta$ (λ a limit ordinal) – axioms of replacement ensure that the operation R is well-defined – and say that s is a stage if $(\exists\alpha)s = R_\alpha$, that x is formed at s if x is subset but not a member of s, and that s is earlier than t if, for some α, β, $s = R_\alpha$, $t = R_\beta$, and $\alpha < \beta$. Then we can prove as theorems of ZF not only the translations into the language of set theory of the axioms of the stage theory, but also those of all those stronger axioms asserting the existence of stages further and further "out" that might have been suggested by the rough description (and those of the instances of the bounding principle which are expressible in $\mathcal{J}$ as well). ZF thus enables us to describe and assert the full first-order content of the iterative conception within the language of set theory.

Although they are not derived from the iterative conception, the reason for adopting the axioms of replacement is quite simple: they have many desirable consequences and (apparently) no undersirable ones. In addition to theorems about the iterative conception, the consequences of replacement include a satisfactory if not ideal[13] theory of infinite numbers,

[12] Worse yet, R_{δ_1} would also seem to be such a model. (δ_1 is the first nonrecursive ordinal.)

[13] An ideal theory would decide the continuum hypothesis, at least.

and a highly desirable result that justifies inductive definitions on well-founded relations.

The axiom of extensionality. The axiom of extensionality enjoys a special epistemological status shared by none of the other axioms of ZF. Were someone to deny another of the axioms of ZF, we would be rather more inclined to suppose, on the basis of his denial alone, that he believed that axiom false than we would if he denied the axiom of extensionality. Although 'there are unmarried bachelors' and 'there are no bachelors' are equally preposterous things to say, if someone were to say the former, he would far more invite the suspicion that he did not mean what he said than someone who said the latter. Similarly, if someone were to say, "there are distinct sets with the same members," he would thereby justify us in thinking his usage nonstandard far more than someone who asserted the denial of some other axiom. Because of this difference, one might be tempted to call the axiom of extensionality "analytic," true by virtue of the meanings of the words contained in it, but not to consider the other axioms analytic.

It has been persuasively argued, by Quine and others, however, that until we have an acceptable explanation of how a sentence (or what it says) can be true in virtue of meanings, we should refrain from calling *anything* analytic. It seems probable, nevertheless, that whatever justification for accepting the axiom of extensionality there may be, it is more likely to resemble the justification for accepting most of the classical examples of analytic sentences, such as 'all bachelors are unmarried' or 'siblings have siblings' than is the justification for accepting the other axioms of set theory. That the concepts of *set* and *being a member of* obey the axiom of extensionality is a far more central feature of our use of them than is the fact that they obey any other axiom. A theory that denied, or even failed to affirm, some of the other axioms of ZF might still be called a set theory, albeit a deviant or fragmentary one. But a theory that did not affirm that the objects with which it dealt were identical if they had the same members would only by charity be called a theory of *sets* alone.

The axiom of choice. One form of the axiom of choice, sometimes called the "multiplicative axiom," is the statement, 'For any x, if x is a set of nonempty disjoint sets (two sets are disjoint if nothing is a member of both), then there is a set, called a *choice set* for x, that contains exactly one member of each of the members of x'.

It seems that, unfortunately, the iterative conception is neutral with respect to the axiom of choice. It is easy to show that, since, as is now known, neither the axiom of choice nor its negation is a theorem of ZF, neither the axiom nor its negation can be derived from the stage theory. Of course the stage theory, which is supposed to formalize the rough description, could be extended so as to decide the axiom. But it seems that no additional axiom, which would decide choice, can be inferred from the rough description without the assumption of the axiom of choice itself, or some equally uncertain principle, in the inference. The difficulty with the axiom of choice is that the decision whether to regard the rough description as implying a principle about sets and stages from which the axiom could be derived is as difficult a decision, because essentially the same decision, as the decision whether to accept the axiom.

Suppose that we tried to derive the axiom by arguing in this manner: Let x be a set of nonempty disjoint sets. x is formed at some stage s. The members of members of x are formed at earlier stages than s. Hence, at s, if not earlier, there is a set formed that contains exactly one member of each member of x. But to assert this is to beg the question. How do we know that such a choice set *is* formed? If a choice set is formed, it is indeed formed at or before s. But how do we know that one is formed at all? To argue that at s we can choose one member from each member of x and so form a choice set for x is also to beg the question: "we *can't* choose" one member from each member of x if there is no choice set for x.

To say this is not to say that the axiom of choice is not both obvious and indispensable. It is only to say that the justification for its acceptance is not to be found in the iterative conception of set.

What is the iterative conception of set?

CHARLES PARSONS

I intend to raise here some questions about what is nowadays called the 'iterative conception of set'. Examination of the literature will show that it is not so clear as it should be what this conception *is*.

Some expositions of the iterative conception rest on a 'genetic' or 'constructive' conception of the existence of sets. An example is the subtle and interesting treatment of Professor Wang.[1] This conception is more metaphysical, and in particular more idealistic, than I would expect most set theorists to be comfortable with. In my discussion I shall raise some difficulties for it.

In the last part of the paper I introduce an alternative based on some hints of Cantor and on the Russellian idea of typical ambiguity. This is not less metaphysical though it is intended to be less idealistic. I see no way to obtain philosophical understanding of set theory while avoiding metaphysics; the only alternative I can see is a positivistic conception of set theory. Perhaps the latter would attract some who agree with the critical part of my argument.

However, the positive part of the paper will concentrate on the notion of proper class and the meaning of unrestricted quantifiers in set theory. That these issues are closely related is evident since in Zermelo-type set theories the universe is a proper class.

The concept of set is also intimately related to that of *ordinal*. Although this relation will be remarked on in several places, a more complete account of it, and thus of the more properly *iterative* aspect of the iterative conception, will have to be postponed until another occasion.

I

One can state in approximately neutral fashion what is essential to the 'iterative' conception: sets form a well-founded hierarchy in which the

Reprinted with the kind permission of the author, the editors and the publisher from *Proceedings of the 5th International Congress of Logic, Methodology and Philosophy of Science 1975, Part I: Logic, Foundations of Mathematics, and Computability Theory*, Robert E. Butts and Jaakko Hintikka, eds., D. Reidel 1977, pp. 335–67.

[1] Wang (1974, chap. 6, which is reprinted in this volume). A widely cited writer whose viewpoint I would also describe as genetic is Shoenfield (1967: 238–40).

elements of a set precede the set itself. In axiomatic set theory, this idea is most directly expressed by the axiom of foundation, which says that any non-empty set has an '$\in$-minimal' element.[2] But what makes it possible to use such an assumption in *motivating* the axioms of set theory is that other evident or persuasive principles of set existence are compatible with it and even suggest it, as is indicated by the von Neumann relative consistency proof for the axiom of foundation.

On the 'genetic' conception that I will discuss shortly, the hierarchy arises because sets are taken to be 'formed' or 'constituted' from previously given objects, sets of individuals [see footnote 2]. But one can speak more abstractly and generally of the elements of a set as being *prior* to the set. In axiomatic set theory with foundation, this receives a mathematically explicit formulation, in which the relation of priority is assumed to be well-founded.

For motivation and justification of set theory, it is important to ask in what this 'priority' consists. However, for the practice of set theory from there on, only the abstract structure of the relation matters. Here we should recall that the hierarchy of sets can be 'linearized' in that each set can be assigned an ordinal as its *rank*. Individuals, and for smoothness of theory the empty set,[3] obtain rank 0. In general, the rank of a set is the least ordinal greater than the ranks of all its elements.[4]

It should be observed that the notion of well-foundedness is prima facie second-order and thus is not totally captured by the first-order

[2] I shall consider throughout set theories which allow individuals (*Urelemente*); this requires trivial modifications of the most usual axioms, but the choice among possible ways of doing this is of no importance for us. In extensionality and foundation, the main parameters are restricted to sets (or at least nonindividuals, if classes are admitted).

The literature on the foundations of set theory does not sufficiently emphasize that the exclusion of individuals in the standard axiomatizations of set theory is a rather artificial step, taken for the convenience of pure mathematics. An applied set theory would normally have to have individuals. What is more relevant for the present discussion is that some of the intuitions about sets with which set theory starts concern sets of individuals. First-order set theory with individuals is compatible with the assumption that there are no individuals and therefore with the usual individual-free set theory.

[3] From the genetic point of view, this is an artifice: individuals are presumably given prior to any sets, even the empty set, so that if rank directly reflects order of construction, the empty set should have rank 1. The same holds for the alternative viewpoint I present below.

[4] In a set theory with individuals, some usual theorems about ranks, for example that for every ordinal α there is a set R_α of all sets of rank $<\alpha$, require the assumption that there is a set of all individuals. It follows that there cannot be too many individuals; for example, ordinals cannot all be construed as individuals. The plausibility of this assumption depends on the intended application. It would be a piece of highly dubious metaphysics to assume there is a set of *absolutely* all individuals, if for no other reason because it is not settled once and for all what *is* an individual. Pure mathematics should be independent of this question; for it the individuals can be an arbitrary set, class, or sometimes structure. However, below I shall assume that the individuals constitute a set.

axiom of foundation.[5] ZF with foundation has models in which the relation representing membership is not well-founded. However, it can be seen that in such a model there is a (not first-order definable) binary relation on the universe for which *replacement* fails. The axioms of separation and replacement are also prima facie second-order, and the fault for such failure of well-foundedness lies in the fact that their full content is not captured by the first-order schemata. But then the problem of stating clearly the iterative conception of set is bound up with the problem of the relation of set theory to second-order conceptions. This problem was already present at the historical beginning of axiomatic set theory with Zermelo's use of the notion of 'definite property'.

The idea that the elements of a set are prior to the set is highly persuasive as an approach to the paradoxes. If we suppose that the elements of a set must be 'given' before the set, then no set can be an element of itself, and there can be no universal set. The reasoning leading to the Russell and Cantor paradoxes is cut off.

However, one does not deal so directly with the Burali-Forti paradox. Why should it not be that all ordinals are individuals and therefore 'prior' to all sets, so that there is no obstacle of this kind to the existence of a set of all ordinals? To be sure, once we look at things in this way it becomes persuasive to view the Burali-Forti argument as just a proof that there is no set of all ordinals. Moreover, the conception of ordinals as order types of well-ordered sets would suggest that for any ordinal there is at least one set of that order type to which it is not prior, so that the existence of a set of all ordinals would imply that later in the priority ordering no new order types could arise. But if there were a set of all ordinals, W, then $W \cup \{W\}$ would have just such an order type.

That ordinals need to fit into a priority ordering with sets and indeed be 'cofinal' with them seems to have been neglected in discussion of iterative set theory, perhaps because in the formal theory ordinals are construed as sets, so that this happens automatically.

One would like to maintain that the requirement of priority is the *only* principle limiting the existence of sets, so that at a given position 'arbitrary multitudes' of objects which are at earlier positions form sets. Although it is difficult to make sense of this, it at least should imply a comprehension principle: given a predicate 'F' which is definitely true or false of each object prior to the position in question, there is *at* the position a set whose elements are just the prior objects satisfying 'F'. In particular, the axiom of separation follows: since the elements of x are prior

[5] However, in set theories with classes foundation for classes follows from its assumption for sets.

to x, $\{z: z \in x \wedge Fz\}$ exists and is not posterior to x. But to apply this idea more generally, some way of marking positions is needed. The genetic approach in effect assumes such to be given. Conversely, if set theory is assumed, the ordinals offer such a marking.

II

We have now gone about as far as we can without explaining what I have called the genetic approach. Put most generally, it supposes that sets are 'formed', 'constructed', or 'collected' from their elements in a succession of stages. The first part of this idea has some plausibility as an interpretation of some of Cantor's preliminary remarks about what a set is. Thus Cantor's famous 'definition' of 1895:

> By a 'set' we understand any collection M into a whole of definite, well-distinguished objects of our intuition or our thought (which will be called the 'elements' of M).[6]

If we were to take 'collection into a whole' quite literally as an operation, then the priority of the elements of a set to the set would simply be priority in order of construction. Cantor's language suggests rather that 'collection' (*Zusammenfassung*) is an operation of the *mind;* in this case the requirement would be that the objects be represented to the *mind* before the operation of collection is performed. However, it will be clear that as it stands this temporal reading is too crude.

It may seem that these notions belong only to the early history of set theory, and in particular that they would have disappeared with the discrediting of logical psychologism at the end of the last century. But the fact is that they are to be found in the contemporary literature. Thus Schoenfield writes (1967: 238):

> A closer examination of the paradox [Russell's] shows that it does not really contradict the intuitive notion of a set. According to this notion, a set A is formed by gathering together certain objects to form a single object, which is the set A. Thus before the set A is formed, we must have available all of the objects which are to be members of A.

Although Schoenfield says that we form sets 'in successive stages', he does not offer an interpretation, temporal or otherwise, of the stages, although he does use 'earlier' to express their order.

[6]Cantor 1932: 282. Wang calls this definition "genetic" (1974: 188) and speculates that the difference between this and previous ones (in particular, Cantor 1932: 204) may be due to Cantor's awareness of the Burali-Forti paradox. It should be remarked that a genetic conception of *ordinals* is intimated in (Cantor 1932: 195–6), a text from 1883.

Wang writes, "The set is a single object formed by collecting the members together" (1974: 181) [530 in this volume]. He recognizes that the concept of collecting is highly problematic and makes an interesting attempt to explain it; he interprets it as an operation of the mind. We shall discuss his views shortly.

We now have the familiar conception of sets as formed in a well-ordered sequence of stages, where a set can be formed at a given stage only from sets formed at earlier stages and from whatever objects were available at the outset.

The language of Cantor, Shoenfield, and Wang invites regarding the intuitive concept of set as analogous to the concepts of constructive mathematics, where one also uses the idea of mathematical objects as constructed in successive stages, and where there is no stage at which all constructions are complete. An immediately obvious limitation of the analogy is that in the typical constructive case (e.g. orthodox intuitionism) the succession of stages is simply succession in *time,* and incompletability arises from the fact that the theory is a theory of an idealized finite mind which is located at some point in time and has available only what it has constructed in the past and its intentional attitudes toward the future. The same interpretation of iterative set theory would require that the stages be thought of as a kind of 'super-time' of a structure richer than can be represented in time on any intelligible account of construction in time. It is hard to see what the conception of an idealized mind is that would fit here; it would differ not only from finite minds but also from the divine mind as conceived in philosophical theology, for the latter is thought of either as in time, and therefore as doing things in an order with the same structure as that in which finite beings operate, or its eternity is interpreted as complete liberation from succession.

It may seem that there is a much more obvious conflict between iterative set theory and a constructive interpretation of it: set theory is the very paradigm of a *platonistic* theory. As is customary in discussing the foundations of mathematics, platonism means here not just accepting abstract entities or universals but epistemological or metaphysical realism with respect to them. Thus a platonistic interpretation of a theory of mathematical objects takes the truth or falsity of statements of the theory, in particular statements of existence, to be objectively determined independently of the possibilities of our knowing this truth or falsity. Contrast, for example, the traditional intuitionist conception of a mathematical statement as an indication of a 'mental construction' that constitutes a proof of the statement.

Perhaps it would be rash to rule out an interpretation of set theory that would not be platonistic.[7] But in any case it seems that a platonistic interpretation is flatly incompatible with viewing the 'formation' of sets as an operation of the mind. However, that there is not a direct contradiction should be evident when we observe that we are concerned in set theory with what formations of sets are *possible*. In contrast to the situation with intuitionism, we do not require that a statement to the effect that it is possible to 'construct' a set satisfying a certain condition should be itself an indication of a construction. Even if we construe the formation of sets as a mental operation, what is possible with respect to such formation can be viewed independently of our knowledge. Thus there is a *prima facie* resolution of the difficulty posed by platonism.

However, we have not reckoned yet with the actual content of the set-theoretic principles that seem to require a platonistic interpretation, such as the combination of classical logic with the postulation of a set of all sets of integers. In an iterative account, the individual steps of iteration are in Wang's word "maximum" (1974: 183 [532 in this volume]. Namely we regard as available at any given stage any set that *could* have been formed earlier. We could represent this assumption as that if a set *can* be formed at a given stage, then it *is* formed (or at least that it *exists* at that stage). This of course has effects on what can be formed later, since every *possibility* of set formation at stage α is such that its result is available at later stages and can therefore enter into further constructions.

We can illustrate this by the manner in which these ideas are used to justify the power set axiom. At a given stage α, any 'multitude' of available objects can be formed into a set. Let x be a set formed at stage α, and let y be a subset of x. Since the elements of y are all elements of x, they must have been available at stage α. Hence y *could have been formed* at stage α. x is available from stage $\alpha+1$, on; from our assumption it follows that y is available as well. Thus at stage $\alpha+1$, *every* subset x is available and $\mathfrak{P}(x)$ can be formed.

We should distinguish two principles that are playing a role here, and which can be confused with one another. One is the 'arbitrary' nature of sets, which, following Wang, we have expressed (provisionally) by saying that any 'multitude' of available objects can be formed into a set. The other is the principle that allows the transition from possibility of formation at stage α to availability at stage $\alpha+1$, perhaps by way of existence at stage α. Both principles may be taken to arise from the idea we

[7]Formalism apart, one ought not to rule out the possibility of an interpretation of set theory along constructivist lines, particularly in view of the broadening of the intuitionist outlook in recent years. ZF has recently been shown consistent relative to some set theories based on intuitionist logic. See Friedman (1973) and Powell (1975).

expressed above that the priority requirement, here interpreted to mean that sets are formed from *available* objects, is the *only* constraint on the existence of sets. But the second principle begins to undercut the idea of sets as *formed* from available objects, since the successive stages of formation are required *only* because a set must be formed from available objects, and not because of any successiveness in the process of formation itself. The question arises whether the interpretation of the priority of the elements of a set to the set in terms of order of construction does not reduce to viewing this priority as a matter of *constitution:* the elements are prior because they *constitute* the set (to use a more abstract phrase than they are its *parts,* which would invite inferences inappropriate to set theory). This view is close to what I advocate below, but it is quite different from the conception of set formation as an operation of the mind.

Wang makes an interesting attempt to develop the latter idea. He says that a multitude can be formed into a set only if its "range of variability" is "in some sense intuitive" (1974: 182) [531 in this volume]. I shall for the time being accept the notion of 'multitude'; the problems concerning it are related to the question of the notion of *class* in set theory. Wang indicates that to form a set is to "look through or run through or collect together" all the objects in the multitude.[8] Thus a condition for a multitude to form a set is that it should be possible thus to 'overview' it. This overviewing is a kind of intuition, presumably analogous to perception. Of course infinite multitudes can be 'overviewed'.

Clearly Wang does not maintain that human beings have the capacity to "run through" infinite collections. He speaks of overviewing "in an *idealized* sense" (1974: 182) [531 in this volume]. In other words, he has a highly abstract conception of the possibilities of intuition. In constructive conceptions of the arbitrary finite, we already disregard the actual bounds of human capacities, in the sense that if a sequence of steps has been performed we always *can* perform a further step, and any operation can be iterated. Wang's idealized overviewing carries such abstraction further in that finitude and even the limitations posed by the continuous structure of space-time (as the setting of the objects of perception and even of the mind itself) are disregarded.

The question arises what force it still has, on Wang's level of abstraction, to treat the possibilities involved in this kind of motivation of the axioms of set theory as possibilities of *intuition*. The analogy with sense-perception which is central to the constructive conception of intuition in Brouwer or to Hilbert's distinction between intuitive and formal mathe-

[8]Wang may be developing the remark of Gödel (1964: 272, no. 40) [484, n. 26, in this volume] that the function of the concept of set is "synthesis" in a sense close to Kant's.

matics seems to be almost totally lost. Consider Wang's remarks on the axioms of separation and power set. The former is stated thus: If a multitude A is included in a set x, then A is a set.

> Since x is a given set, we can run through all members of x, and, therefore, we can do so with arbitrary omissions. In particular, we can in an idealized sense check against A and delete only those members of x which are not in A. In this way, we obtain an overview of all the objects in A and recognize A as a set (1974: 184) [533 in this volume].

The idealization seems to include something like omniscience: A may be given in some way that does not independently of the axiom assure us that it is a set, and yet we can use it in order to 'choose' the members of x that are in A. A may of course be given to us by a predicate containing quantifiers that do not range over a set; in deciding whether an element y of x is to be deleted, we cannot 'run through' the values of the bound variables as part of the process of checking y against A. It is not clear what more structured account of 'idealized checking' would yield the result Wang needs. An alternative would be to view subsets as run through not by verification but by arbitrary selection. But if a predicate is given, how are we to 'select' just those elements of x that satisfy it unless we can decide which ones do?[9]

More strain on the concept of intuition appears in Wang's treatment of the power set axiom (1974: 184) [534 in this volume]:

> We have $\cdots$ an intuitive idea of running through with omissions. This general notion $\cdots$ provides us with an overview of all cases of AS [separation] as applied to x.

By saying that we have an "intuitive idea" of running through with

[9] Cf. the fact that in intuitionism the classical notion of set splits into those of spread and species.

It is of interest to look at a case, namely the hereditarily finite sets, where something like arithmetical intuition does yield the axiom of separation. Here we can argue by induction on n that there is a w such that

$$(*) \qquad (\forall z)(z \in w \leftrightarrow z \in \{x_1 \cdots x_n\} \wedge Fz).$$

For if $n=0$, $w=\Lambda$. Suppose w satisfies (*) and consider $\{x_1 \cdots x_{n+1}\}$. If $\neg Fx_{n+1}$ then w satisfies

$$(\forall z)(z \in w \leftrightarrow z \in \{x_1 \cdots x_{n+1}\} \wedge Fz);$$

if Fx_{n+1}, then $w \cup \{x_{n+1}\}$ satisfies the condition. In effect this argument shows us that of the possible subsets of a finite set *one* satisfies the separation condition, without telling us which one.

One might consider interpreting the requirement that a set x can be 'overviewed' as meaning that it can be run through in a well-ordered way and then attempt a transfinite analogue of this inductive argument. It seems that to handle limit cases such an argument requires replacement, and of course separation can be deduced from replacement in a more trivial way without assuming x to be well-orderable.

omissions, he does not only mean that a *case* of such running through is intuitable, for that would not yield the result. Rather, for a given set x the concept of such runnings through is intuitive in the sense that 'we can' run through *all* cases of it. Something of the content of the idea of intuitive running through seems to be lost here. Clearly in the case of small finite sets of manageable objects, we really do see all the elements 'as a unity' in a way that preserves the articulation of the individual elements. Somewhat larger sets can be seen by a completable succession of steps of bringing one (or a few) objects under one's purview. If we consider arbitrary iterations of such steps, there is no longer any limit to how large a (finite) set can be thus intuited. We also have a simple and clear generative rule for sets such as the natural numbers, though the process of sensibly intuiting is in this case incompletable, so that the *givenness* to us of such a set depends in a more essential way on conception. But regarding the natural numbers as intuitable as a whole amounts just to abstracting from the above incompletability. There is another qualitiative leap in dealing with all sets of integers, as has often been remarked on in the literature (as indeed Wang himself emphasizes when he stresses the importance of impredicativity).[10] Here, however we understand the notion of an 'arbitrary set' of integers, say by some picture of arbitrary selection, we do not have the conceptual grasp of what the totality contains that would be given by some method of generating them. The divorce from *sensible* intuition involved in treating this totality as 'intuitable' seems complete, unless perception is used only as a source of quite remote analogies. Two mathematical symptoms of this situation are the absence of a definable well-ordering of the continuum and our inability to solve the continuum problem.

I ought to make clear that I understand by 'intuition' a quasi-perceptual manner in which an object is presented to the mind. In this I follow Kant. The word 'intuition' is also used in the philosophy of mathematics and otherwise for any manner by which propositions can be known where this knowledge is not largely accounted for by deductive or inductive reasoning. There is a tendency to confuse these two senses. As for the appropriateness of my sense of intuition to Wang, I should point out that the other concept of intuition does not distinguish sets from 'multitudes' or other primitive notions that might enter into evident set-theoretic axioms. Moreover, intuition in the latter sense is purely *de dicto*, intuition *that* certain propositions are true, while Wang clearly requires intuition *de re*, intuition *of* sets.[11]

[10] For example the classic formulation of Bernays (1935: 275–6) [259–60 in this volume]. Cf. Wang (1974: 183) [532 in this volume].

[11] This distinction is explicitly made by Steiner (1975: 130–1). However, Steiner regards

I no longer understand Wang's talk of 'intuitively running through' where it is applied to the set of all sets of integers. In the above I have perhaps connected intuition more closely with the senses (more abstractly Kant's "sensibility") than Wang would find acceptable. But even quite abstract marks of sensibility, such as the structure of time, are lost in this case.

However, there might be an interpretation on which Wang's hypothesis that a 'multitude' is a set if and only if it is an object of intuition would be logical and ontological rather than perceptual and epistemic. To be an object of intuition would be simply to be an object rather than a Fregean "concept" or perhaps a property. In other words, Kant's contrast of intuitions as "singular representations" (*Logic*, §1) would give virtually the only essential mark of intuition.[12] Although this interpretation would bring Wang's hypothesis into accord with the views I express below, I shall not pursue it further in this paper.

I want now to turn to the axiom of replacement, about which Wang has most interesting things to say. Wang writes:

> Once we adopt the viewpoint that we can in an idealized sense run through all members of a given set, the justification of *SAR*[13] is immediate. That is, if, for

only *de dicto* mathematical intuition as defensible. I have sketched in previous writing an account of arithmetical intuition on Kantian lines (1965: 201–3; 1969; 1971: 158–62, 166–7). Curiously, Steiner cites me and then says, "No one today, however, upholds hard-core intuition – the direct intuition of mathematical *objects*" (*ibid.*), although he then mentions Gödel as a possible exception. Since I was trying to elucidate the *forward* character of Kantian intuition, perhaps Steiner did not consider arithmetical intuition on my view to be "direct intuition of mathematical *objects*", particularly in (1971).

Some comment is in order on Gödel's view of mathematical intuition, particularly since he explicitly says, "We do have something like a perception also of the objects of set theory" (1964: 270) [483–4 in this volume]. This seems to commit Gödel to intuition *de re*. His immediately following remark does not give any argument for this; he says only that it "is seen from the fact that the axioms force themselves on us as being true", which implies only intuition *de dicto*.

However, it seems clear that Gödel holds (1964: 271 bottom) [484 in this volume], that our ideas of objects of certain kinds contain "constituents" which are *given* (not "created" by thinking) on the basis of which we "form our ideas" of these objects and postulate theories of them. "Evidently the 'given' underlying mathematics is closely related to the abstract elements contained in our empirical ideas" (1964: 272) [484 in this volume]. In the case of set theory, Gödel does not give any indication of wanting to distinguish sets as objects of intuition from other entities (such as "properties of sets" (1964: 264 n. 18) [476 n. 15 in this volume]) that the axioms might refer to.

Elsewhere Gödel, in contrast to the above passage, contrasts the intuitive with the abstract (1958: 281). There he seems to be using 'intuition' in a much narrower and more strictly Kantian sense. Of course there he is writing in German; possibly he would not use 'Anschauung' in the sense in which he uses 'intuition' in (1964).

[12]According to Hintikka, this is Kant's own view. See Hintikka 1969a and other essays reprinted in 1973 and 1974.

[13]SAR is the statement: if b is an operation and b_x is a set for every member x of a set y, then all these sets b_x form a set (Wang 1974: 186) [535 in this volume].

> each element of the set, we put some other given object there, we are able to run through the resulting multitude as well. In this manner, we are justified in forming new sets by arbitrary replacements. If, however, one does not have this idea of running through all members of a given set, the justification of the replacement axiom is more complex. (1974: 186) [536 in this volume]

The picture here is marvelously persuasive; for me, it expresses very well why the axiom of replacement seems obvious. But something like the omniscience assumption of the discussion of separation is present in the remark that "*we put* some other given object there" and "are able to run through the resulting multitude." What is much more revealing is that the objects seem to have no relevant internal structure: Our ability to 'run through' a multitude is preserved if we replace its elements by *any* other objects, for example by much larger sets. It is as if the objects were given only as wholes, or at least that any internal structure would not affect the possibility of running through the totality. A model for this (conceptual rather than intuitive) is the case where the objects are given only by names. Wang seems to be making an hypothesis here, although I do not feel the same qualms as in the case of the power set about taking it as an hypothesis about what is intuitable. It is of course the *combination* of the replacement with the power set axiom that yields sets of very high ranks.

Wang expresses by his picture the idea, present in the earliest intimation of the axiom of replacement (Cantor 1932: 444; cf. Wang 1974: 211 [562 in this volume]) that whether a multitude forms a set depends only on its cardinality and not on the 'internal constitution' or relations of its elements. Put in this way, the axiom is not a principle of *iteration* of set formation, in line with the conclusion of Boolos (1971: 228–9 [500–1 in this volume]) that it does not follow from the iterative conception of set.[14] The most direct justification of replacement by appeal to ideas about stages seems to me somewhat circular.[15] That of Gödel cited by

[14]Boolos seems, however, to arrive at his conclusion too easily. He seems to assume that the 'stages' and their ordering have to be given independently of the concept of set, at least for the expression of *the* iterative conception. His actual axioms about stages (and a further possible one he mentions on p. 227 [499 in this volume]) would permit the stages to be ordered by a very simple recursive well-ordering. It seems to me that sets and 'stages' ought to be 'formed' together, so that the formation of certain sets should make possible going on to further stages.

However, the most obvious principle of this kind, that if a well-ordering has been constructed then there is a stage such that the earlier stages are ordered isomorphically to the given well-ordering, is weaker than the axiom of replacement.

[15]Thus Shoenfield (1967: 240) deduces replacement from a "cofinality principle": if "we have a set A, and ⋯ we have assigned a stage S_a to each element a of A.... There is to be a stage which follows all of the stages S_a" (p. 239). However, he justifies this by saying, "Since we can visualize the collection A as a single object (viz., the set A), we can also visualize the collection of stages S_a as a single object; so we can visualize a situation in which all

Wang (1974: 186 and 221, n. 5) [536 and n.4, 536 in this volume] I find less immediate and persuasive.

Although I admit that Wang's picture (apart from the question of omniscience) offers a plausible hypothesis about what is intuitable,[16] it seems to me to be equally plausible as an hypothesis about what can be thought or about what can *be*, and the latter interpretations fit better the case of power set. I want now to pursue the genetic conception of sets in this direction.

III

In the preceding section we saw a number of difficulties with the idea that sets are 'formed' from their elements, in particular by an activity of running through in intuition. I want now to suggest a more 'ontological' view of the hierarchy of sets.

The earliest attempt that we know to explain the paradoxes of set theory and to develop set theory in a way that avoids them is in Cantor's famous letter to Dedekind of July 28, 1899 (1932: 443–7). Cantor there presupposes his earlier 'many into one' characterizations of the notion of set, such as that of 1895 cited above. He begins (p. 443) with "the concept of a definite multiplicity (*Vielheit*)" What he calls an *inconsistent* multiplicity is one such that "the assumption of a 'being together' (*Zusammensein*) of *all* its elements leads to a contradiction". A consistent multiplicity or set is one whose "being collected together to 'one thing' is possible". It is noteworthy that Cantor here identifies the possibility of all the elements of a multiplicity ***being together*** with the possibility of their being collected together into one thing. This intimates the more recent conception that a 'multiplicity' that does not constitute a set is *merely potential*, according to which one can distinguish potential from actual being in some way so that it is impossible that *all* the elements of an inconsistent multiplicity should be actual.

I am here interpreting Cantor to mean that where there is an essential obstacle to a multiplicity's being collected into a unity, this is due to the fact that in a certain sense the multiplicity does not exist. It does not exist

the stages are completed" (*ibid.*). Here he is assuming that "visualizability as a single object" is preserved by replacement of a by S_a; but that is just the principle of replacement. Wang's picture seems more fundamental than the kind of argument Shoenfield gives.

Wang gives a similar argument (1974: 220 n. 4) [535–6 n.3 in this volume].

The argument does obtain general replacement from the special case where the range of the replacing function consists of stages.

[16]This plausibility is perhaps reinforced by the fact that replacement holds for the hereditarily finite and the hereditarily countable sets.

as a totality of its elements; if it did, they would form a unity or could at least be *collected* into a unity. But in the case of inconsistent multiplicities, this is impossible. The sense of this non-existence needs some further elucidation which Cantor does not supply. The language of potentiality and actuality is not in the text, though Cantor may have been suggesting it in calling an inconsistent multiplicity *absolutely infinite* (p. 443).

What seems to me of interest in the present connection in these hints of Cantor is that he seems to be trying to distinguish sets from "inconsistent multiplicities" without real use of any metaphor of *process* according to which sets are those multiplicities whose 'formation' can be 'completed'.[17] Such a metaphor makes the idea of an inconsistent multiplicity as a merely potential totality rather easy. I suggest interpreting Cantor by means of a modal language with quantifiers, where within a modal operator a quantifier always ranges over a set (not, however, one that is explicitly given or even that exists in the 'possible world' it might be taken to range over). Then it is not possible that all elements of, say, Russell's class exist, although for any element, it is possible that *it* exists. As it stands this conception requires it to be meaningful to talk of *any* set (or any object), even though the range of *this* quantifier does not constitute a unity; the elements of its range cannot all 'exist together'. However, at least some such talk can be replaced by ordinary quantification behind necessity.

What one would like to obtain from this conception is some interpretation of the stages of the iterative conception that also does not depend on the metaphor of process. However, I intend first to look at Cantor's conception of a multiplicity. Wang seems to use "multitude" in the same sense, although he does not use it to translate Cantor's *Vielheit* when he discusses Cantor's 1899 correspondence with Dedekind (1974: 211) [562 in this volume). These notions are among a number which occur in the literature on logic and set theory and which purport to be more comprehensive than the notion of set. The most respectable of these notions is that of (proper) class. We should also mention Frege's "concept", Zermelo's "definite property", and Shoenfield's "collection". Gödel's "property of sets" (1964, n. 18, p. 264) [n. 15, p. 476 in this volume] presumably also belongs on this list.

[17]However, this is not to say that Cantor now conceives sets as having no intrinsic relation to the mind. Hao Wang has pointed out to me that this would be questionable. For example he characterizes a consistent multiplicity as one the totality of whose elements "can be thought of without contradiction as 'being together', so that their being collected together (***Zusammengefasstwerden***) to 'one thing' is possible".

Of all these notions, perhaps the most developed from the philosophical side is Frege's notion of a concept; I shall use it for purposes of comparison. What is then striking is that neither Cantor's nor Wang's notion seems to be derived from predication as Frege's is. Since Cantor's notion is one of the prototypes of the notion of proper class, this fact seems to clash with the actual use of the notion of class in set theory (perhaps with exceptions; see below) according to which classes are derived from predication; Zermelo's "definite property" is a more immediate prototype than Cantor's *Vielheit*. I myself have suggested (1974c) that sets are not derived from predication while classes are.

Cantor in 1899 apparently thought of sets as a species of the genus multiplicity, and then perhaps the non-predicative (if not impredicative!) character of multiplicities in general was needed in order to preserve the 'arbitrariness' of sets against its being restricted by what we might express in language. Frege seems to have obtained the same freedom by his realism about concepts. However, for Frege the nature of a concept could apparently only be explained by appeal to predication (more generally, to 'unsaturated' expressions). The sharp distinction between concepts and objects is a shadow of the syntactical difference between expressions with and without argument places. This difference is then 'inherited' by concepts that are not denoted by expressions of any language we use or understand.

I want to suggest that predication plays a constitutive role in the explanation of Cantor's notion of multiplicity as well and that at least an "inconsistent multiplicity" must resemble a Fregean concept in not being straightforwardly an object. In the Cantorian context, predication seems to be essential in explaining how a multiplicity can be given to us not as a unity, that is as a set. Much the clearest case of this is understanding a predicate. Understanding 'x is an ordinal' is a kind of consciousness or knowledge of ordinals that does not so far 'take them as one' in such a way that they constitute an object. We might abstract from language and speak with Kant of knowledge through concepts, but whatever we make of this the predicational structure is still present.

The philosophy of Kant might suggest another way in which a multiplicity might be given not as a unity, namely as an 'unsynthesized manifold'. It seems clear that in the cases Kant actually envisaged, the objects involved would have the definiteness necessary to constitute a Cantorian 'multiplicity' only if they are a set. Even if we generalize the notion in some way, I do not see how such a 'manifold' can be taken up into explicit consciousness except perceptually (intuitively) or conceptually.

The idea that to be an *object* and to be a *unity* are the same thing is very tempting and has deep roots in the history of philosophy. An object

is something whose identity with itself (represented in different ways) and difference from other objects can be meaningfully talked about; it is then subject to at least rudimentary application of *number*. This line of reasoning inclines us to identify Cantor's "multiplicity" with Frege's concept at least in that a multiplicity which is not a set is not an object. Some such assumption seems necessary to cut off the question why there are not multiplicities whose elements are not sets or individuals: multiplicities are multiplicities of *objects*, and under that condition there are no restrictions on the existence of multiplicities (although possibly on the use of quantifiers over them), but if a multiplicity is an object, then it is a set.

However, we have to deal with the fact that in Frege the gulf between concepts and objects comes from the structure of predication itself, so that a concept is irremediably not an object, even if only one object falls under it. Cantor evidently holds that some multiplicities just *are* sets, in particular those that are not too large. This may seem not a very essential difference: if a concept F is such that there is a set y of all x such that Fx, then the distinction between F and y is just the distinction between $(\ ) \in y$ and y.[18] For an inconsistent multiplicity there is no such 'reducibility'. In view of Russell's paradox, the idea of the predicative nature of the concept will motivate the idea that there should *be* inconsistent multiplicities, but it does not seem to motivate Cantor's particular principles as to what multiplicities are 'consistent'. For reasons which will become clear later, I do not think we have yet captured the sense in which an inconsistent multiplicity is not an object.

Let us look for a moment at the well-known difficulties of Frege's theory of concepts. The conception has the great attraction that it enables us to generalize predicate places without introducing nominalized predicates that purport to denote objects (classes or attributes) – something that has to be restricted on pain of Russell's paradox. But the temptation to nominalize is irresistible, as Frege himself discovered on two fronts. His construction of mathematics required an 'official' nominalization in postulating extensions. 'Unofficial' nominalizations cropped up repeatedly in his own informal talk about concepts and gave rise to the paradox that the concept *horse* is an object, not a concept.[19] At the end of his life Frege decided the temptation was to be resisted and that neither the expression "the concept F" nor the expression "the extension

[18] In a sense $(\ ) \in y$ *is* F, since coextensiveness for concepts is the analogue of identity for objects, but we cannot say that the concepts are *identical*.

[19] Frege 1892a. Only it might be a concept after all, since "is a concept" is syntactically such that it takes *object*-names as subjects, and is therefore a predicate of objects.

Of course a voluminous literature has grown up on this question.

of the concept F" really denotes anything.[20] However, perhaps yielding to temptation can in one way or another be legitimized, even where the extensions postulated are not sets (cf. my, 1974c).

Does the Cantorian concept of "multiplicity" have to be understood realistically? To the extent that *sets* are understood realistically, of course "consistent multiplicities" are mind-independent in the corresponding sense. However, obviously it does not follow from the fact that we allow classical logic and impredicative reasoning about sets that we have to allow either about classes or other more general entities. The suggestion made below that such entities are at bottom *intensions* would imply, if we think of an intension in the traditional way as a meaning entertained by, and in some sense constructed by, the mind, that realism about them is inappropriate. However, in view of the interest of impredicative conceptions of classes for large cardinals, both predicative and impredicative conceptions should be pursued.[21]

[20]Frege (1969–76, 1: 288–9), a text written in 1924 or 1925. Cf. (1969–76, 1: 276–7), from 1919. The late evolution of Frege's thought on these matters is discussed in my (1976).

One can question whether the problem of generalizing predicate places is really solved by Frege's approach. Once we have generalizable variables in predicate places, we have new predicates that are not generalized by the variables in question – predicates which in Frege's semantics denote "second-level concepts". Hence the urge to extend the language by nominalization appears in Frege's context in another form. An ultimate Fregean canonical language would have to be a predicate calculus of order ω of which the semantics can no longer be expressed, unless we admit predicates of infinitely many arguments of different types. Surely *we* understand such a language by a means which from this Fregean point of view is a falsification, namely by a recursion in which *in general* variables with argument places of given types range over relations of these arguments, and each type is reached by finite iteration of the ascent from arguments to function. That involves a 'unification of universes' that Frege rejected, and which essentially contains nominalization.

Frege's logic contained bound variables only for objects and first-level functions, and free variables for one type of second-level function. He refers informally in at least one place to a third-level function (1893–1903, 1: 41), which would seem to be required by the *semantics* of his system. Formally, he thought higher levels dispensable because second-level functions could be replaced by first-level functions in which the function arguments were replaced by their *Wertverläufe* (1893–1903, 1: 42). This was untenable because it depended on the inconsistent axiom V. But of course in a less absolute way to replace functions by sets which are objects is just the procedure of set theory, which then does dispense with 'higher level functions' for most purposes. It is only quite recently, with the discussion of measurable and other very large cardinals, that higher than second-order concepts relative to the universe of sets have had any real application. See especially Reinhardt (1974a) and Wang (1977).

[21]Analogously to the theory of predicatively definable sets of natural numbers, one can explore mathematically the predicative definability of classes relative to the universe of sets. See Moschovakis (1971).

Wang's discussion of the axioms of separation and power set could lead one to think that impredicative reasoning about 'multitudes' is already involved in motivating the axiom of power set. Although this may be psychologically natural, what the power set axiom says is that given a set x there is a set of all sub*sets* of x, not that there is a set of all 'multitudes' whose elements are elements of x. Thus being an arbitrary subset of x has to be definite, but the 'multitude' of them is defined without quantifying over arbitrary *multitudes*. The

Let us now return to Cantor's suggestion that the elements of an "inconsistent multiplicity" cannot all exist together. I do not conclude that an inconsistent multiplicity does not exist in any sense; even the hypothesis that it is not an object will have to be qualified. However, one implication is clear: it is not a totality of its elements; it is not 'constituted' in a definite way by its elements. Its existence cannot require the prior existence of all its elements, because there is no such prior existence.

I wish to explicate the difference between sets and classes by means of some intensional principles about them. From the idea that a set is constituted by its elements, it is reasonable to conclude that it is *essential* to a set to have just the elements that it has and that the *existence* of a set requires that of each of its elements. Exactly how one states these principles depends on how one treats existence in modal languages. I shall assume that the truth of $x \in y$ requires that y exists (Ey). Then we have:

(1) $$x \in y \rightarrow Ex \wedge Ey$$

(2) $$x \in y \rightarrow \Box(Ey \rightarrow x \in y)$$

(3) $$x \notin y \wedge Ey \rightarrow \Box(x \notin y).$$ [22]

My proposal is that these principles should fail in some way if y is an "inconsistent multiplicity" or proper class. Indeed Reinhardt has suggested that proper classes differ from sets in that under counterfactual conditions they might have different elements (1974a: 196). I am endorsing this suggestion as an explication of the intuitions about "inconsistent multiplicities" that I have been discussing.

axiom of separation tells us that *any* 'multitude' of elements of x is a subset of x, so that the 'definiteness' of the property of being a subset of x implies that of being a submultitude of x. But we do not need to assume the definiteness of the latter property; indeed if we think of 'multitudes' intensionally (see below), it is only their *extensions* that become a definite totality by this reasoning.

[22]The most natural and elementary application of these principles is in relation to sets of ordinary objects that are the extensions of predicates contingently true of them. I intend to discuss these matters in a paper in preparation; cf. (1974d) and Tharp (1975).

(1)–(3) exactly parallel familiar principles of identity except that identity is usually treated as independent of existence.

(2) implies that set abstracts are not rigid designators. If 'F' is a predicate that holds of an object x, but not necessarily so, then $\Box(E\{z: Fz\} \rightarrow x \in \{z: Fz\})$ is true with the scope of the abstract outside the modal operator but false with the scope within. I assume that in any possible world $\{z: Fz\}$ is the set of existent z such that Fz in that world.

The free variables in (1)–(3) range over all possible objects, although for the present discussion the appropriate modal logic has *bound* variables ranging only over existing objects. If this treatment of free variables is thought to be too Meinongian, then (3) needs to be replaced by a schema

$$\forall x \Box(x \in y \rightarrow Fx) \rightarrow \Box \forall x(x \in y \rightarrow Fx),$$

or, in the second-order case, by the corresponding second-order axiom.

As before, for a given 'possible world' we should think of the bound variables as ranging over a set, perhaps an R_α (see note 4); but the sets that exist in that world are elements of the domain, while classes are arbitrary subsets of the domain.

Reinhardt does not use an intensional language; in his formulation the actual world V is part of a counterfactual 'projected universe' which is the domain of the quantifiers. He assumes a mapping j on $\mathfrak{P}(V)$ such that if $x \in V$, $jx = x$. We can think of jx as a 'counterpart' of x in the projected universe, in the sense of Lewis (1968). Thus sets are their own counterparts and can be strictly reidentified in alternative possible worlds. A set y can have no new elements in the projected universe and its only new non-elements are all x such that $x \notin V$. This accords with (1)–(3).

For a class P, jP can have additional elements in the projected universe, so that it violates (3), although jP agrees with P for 'actual' objects (elements of V).[23] Reinhardt's extensional language must distinguish jP from P; hence my thinking of jP as a 'counterpart'. A complication is that P itself occurs in the projected universe, though now as a set. Reinhardt himself suggests an alternative reading, by which a class x is an *intension*, so that P in the actual world and jP in the projected universe are the 'values' in two different possible worlds of the same intension. A formal language in which this reading might be formulated is the second-order modal language of Montague (1970), where the first-order variables range over objects and are interpreted (in the manner usual for modal logic) rigidly across possible worlds, and the second-order variables range over intensions, which in the semantics are functions from possible worlds to extensions of the appropriate type; but see pages [523–4].[24]

[23] Thus the relation of P and jP does not contradict (2). However, this is due to the special nature of the projected universe: j is an elementary embedding of the (sets and classes of) the actual world into it. (2) is presumably not an appropriate general principle about proper classes.

The explicit application to set theory of a modal conception of mathematical existence and the use of modal quantificational logic to explicate it seem to originate with Putnam (1967a) [reprinted in this volume]. Putnam does not address the questions about 'trans-world identification' of sets that our principles (1)–(3) are meant to answer. However, it appears that his suggested translation of statements containing unrestricted quantifiers (p. 21) [310 in this volume] requires that a "standard concrete model of Zermelo set theory" should have a structure that is rigid, that is the relations are not changed when considered with respect to an alternative possible world. If this assumption is made, equivalents of (2) and (3) follow from the fact that a standard model is maximal for the ranks it contains (p. 20) [309 in this volume]. On "concreteness", cf. Parsons (1980, footnote 33).

Putnam seems to envisage a first-order formulation, which requires his "models" to be objects. The second-order formulation seems to us more appropriate not only for the set-class distinction but also for explicating the priority of the elements of a set to the set (Section V).

[24] A reformulation of Reinhardt's ideas in an intensional language would be desirable, in particular in order to eliminate the Meinongian ontology of possible non-actual sets and

No doubt what is most interesting about Reinhardt's idea is the impredicative use of proper classes that he combines with it, with the result that the ordinals in V are, in the projected universe, a measurable cardinal (1974b: 22; or Wang 1977: 327). However, in discussing the idea of proper classes as intensions I want to keep the predicative interpretation in mind as well.

IV

Cantor's conception suggests a more radical view than we have drawn from it so far, namely that one can in a sense not meaningfully quantify over absolutely all sets. In (1974a) and (1974c) I sketched all too briefly a 'relativistic' conception of quantifiers in set theory. The idea was that an interpretation that assigns to a sentence of set theory a definite sense would take its quantifiers to range over a set (presumably R_α for some large α), but that normally such a sentence would be so used as to be 'systematically ambiguous' as to *what* set the quantifiers ranged over.[25] A

classes that he uses, especially in (1974b). Montague's intensional logic would not be adequate as it stands for this purpose, since his first-order quantifiers range over all possible objects; however, there is no difficulty in reformulating it to fit an interpretation in which bound variables range over existing objects. If the only alternative possible worlds one wants to consider are those with more ranks, than the version of quantified modal logic of Schütte (1968) is applicable. This has the additional advantage that free variables also range over existing objects.

Pure modal logic, however, would not suffice to state Reinhardt's schema (S4) (1974a: 196), since it expresses a condition on a single 'possible world' for infinitely many formulae.

The question arises how a class P can recur 'in extension' in another possible world such as Reinhardt's projected universe. The answer is that it would be represented by its "rigidification", that is an attribute Q satisfying the condition.

$$\forall x(Px \leftrightarrow Qx) \wedge \Box[\forall x(\Box Qx \vee \Box \neg Qx) \wedge \forall R[\forall x \Box(Qx \rightarrow Rx) \rightarrow \Box \forall x(Qx \rightarrow Rx)]]$$

I assume here that bound variables range over existing objects; otherwise Barcan's axiom would hold and the third conjunct would be unnecessary.

Wang (1977) formulates Reinhardt's ideas in the opposite direction, by eliminating the intensional motivation and thinking of V not as the 'actual world' but as a set which is an 'approximation' to the universe. What *mathematical* interest an intensional formulation would have is not clear; perhaps it would suggest 'intuitionistic' approaches to strong reflection axioms.

(*Added in proof.*) The statement that every attribute has a rigidification is just the formulation appropriate to this setting (without the Barcan formula) of the axiom E^σ (for $\sigma = (e)$) of Gallin (1975: 78). That the second conjunct above is equivalent to $\forall x(\Box Qx \vee \Box \neg Qx)$ follows from Barcan's axiom. However, (1)–(3) (p. 519) and the comprehension axiom of p. 527 are inconsistent with the Barcan formula. Axiomatizations of set theory based on the ideas of this paper are described in Parsons 1981.

[25] Such a conception is hinted at in Zermelo (1930); see especially p. 47.

The conception of quantification over all sets advanced here is close to that of Putnam (1967a) [reprinted in this volume], except for the addition of the concept of systematic ambiguity.

merit of this view was (1974c: 8, 10–11) that it yielded a kind of reduction of classes to sets. Since I have here followed Cantor and Reinhardt in viewing classes as quite different from sets, here I can defend this relativistic position only in a modified form. I shall now present my understanding of the matter.

What *does* follow from the thesis that the elements of an inconsistent multiplicity cannot all exist together is that quantification over all sets does not obey the classical correspondence theory of truth. The totality of sets is not 'there' to constitute any 'fact' by virtue of which a sentence involving quantifiers over all sets would be true. The usual model-theoretic conception of logical validity thus leaves out the 'absolute' reading of quantifiers in set theory.

If the only constraint on an interpretation of a discourse in the language of set theory is that it should make statements proved in first-order logic from axioms accepted by the interpreter *true*, then the interpreter needs only a minimally stronger theory than that applied in the discourse to interpret it so that the quantifiers range over some R_α.[26] However, it seems clear that this condition is too weak, and it remains so even if the interpreter seeks to capture not just one particular discourse but what the set theorist he is interpreting might be taken to be *disposed* to assent to. This is the case envisaged in (1974c: 10), where we suppose that the interpreter takes the quantifiers to range over R_α for an α with an inaccessibility property undreamed of by the speaker.

Let us suppose this inaccessibility property to be P and that the interpreter chooses the least such α. One weakness of the reading is that it takes $(\exists\alpha)P\alpha$ to be false; although the supposition that the speaker is not disposed to assent to it is reasonable, the result is arbitrary in that we have no reason to suppose the speaker disposed to dissent from it.

A more decisive objection is that if the interpreter gets the speaker to understand P and convinces him that there is a cardinal satisfying it, then he must attribute to the speaker a meaning change brought about by this persuasion: previously his 'concept of set' excluded P-cardinals; now it admits them.

The speaker, however, can (going outside the language of set theory and talking of himself and his intentions) question this and say that P-cardinals are cardinals in just the sense in which he previously talked of cardinals; he will presumably reinforce this by assenting to a number

[26]The existence of an R_α such that V is an elementary extension of R_α is provable in ZF plus impredicative classes (Bernays-Morse set theory; see Drake (1974: 125)). What is essential, however, is not impredicative classes but allowing bound class variables in instances of replacement; one could use the system NB^+ mentioned in my (1974a, note 15).

of statements that follow directly from the existence of a P-cardinal by axioms or theorems of set theory he accepted previously.

The idea that quantifiers in set theory are systematically ambiguous was meant to meet this kind of objection by saying that the interpretation of the speaker's quantifiers as ranging over a single R_α cannot be an exactly correct interpretation, since it fixes the sense of statements whose sense is not fixed by their use to this degree. However, it still seems to imply that the speaker who is convinced of the existence of a P-cardinal undergoes a meaning change in a weaker sense, in that the ambiguity of his quantifiers is reduced by 'raising the ante' as to what degree of inaccessibility an α has to have so that R_α will 'do'.

We should observe that assertions in pure mathematics are made with a presumption of *necessity;* if we attribute this to our speaker we can see how P-cardinals are immediately captured by his previous set theory, since necessary generalizations are not limited in their force to what 'actually' exists. We can see the 'meaning change' in accepting P-cardinals as analogous to the speaker's considering a different possible world or range of possible worlds.

The force of this analogy is limited, as we can see by a little further reflection on the conception of 'inconsistent multiplicities' as intensions. It seems that we cannot consider a proper class as given even by an *intension* that is definite in the sense of, say, possible-world semantics, as a function from possible worlds to extensions. To begin with, it is only by an interpretation external to a discourse that one can speak in full generality of the range of its quantifiers and the extensions of its predicates. The systematic ambiguity of the language of set theory arises from the fact that such an interpretation can itself be mapped back into the language of set theory when stronger assumptions are made. Thus we should think of predicates whose 'extensions' are proper classes as really not having *fixed* extensions.[27]

This situation does not change if we enlarge the language of set theory to an intensional language. Here we are able to express the 'potentiality

[27]Cf. the remarks on the discomfort evinced by use of proper classes in Wang (1977, footnote 8). However, Wang does not make clear whether this discomfort would be removed if we confine ourselves to thinking "of these large classes as extensions or ranges of properties".

The point which I would emphasize is that if the language of set theory with quantifiers read as ranging over 'all sets' has a 'fixed' or 'definite' sense, then it is naturally extended by a satisfaction predicate, and definiteness of sense is preserved. But in the extended language one can of course construe the classes required by the Bernays-Gödel theory. Iteration of the procedure yields more classes.

In Wang's terms, this justification of classes no doubt falls within the conception of them as "extensions or ranges of properties". Still, such an enlargement of the language of set of the totality of sets' in that it is necessarily true that the domain of a bound variable *possibly* exists as a set. But however such an intensional language is formulated, it will still be possible to read it in a set-theoretic possible world semantics, and even if on the most straightforward reading the union of the domains for all possible worlds is all sets, the assumption that there is a set that realizes the properties of this union presumably has the same plausibility that other such reflection principles have. In such a model, of course a second-order intension will be represented by a set.[28]

We should not be surprised at this; it is really a consequence of the general nature of true 'systematic ambiguity', where there is no general concept of 'possible interpretation' which is not either inadequate or infected with the same difficulties as the language it interprets. Otherwise one could resolve the ambiguity as generality (meaning by A, 'A on every possible interpretation') or indexically, by some contextual device or convention indicating which interpretation is meant. Russell's "typical ambiguity" was essential in that according to his theory of meaning there was no way of expressing by a single generalization *all* instances of a formula where the variables were understood as typically ambiguous. In (1974b) I handled semantical paradoxes by observing that paradoxical sentences could be taken to have no truth-value or not to express propositions on the interpretations presupposed by the semantic concepts occurring in them, while obtaining definite truth-values or coming to express propositions on interpretations 'from outside'. But at some point there must on this account be systematic ambiguity, or else one could generate 'super' paradoxes such as 'this sentence is not true on any interpretation'.

Thus although it is true to say that a proper class is given to us only 'in intension', this statement does not have quite its ordinary meaning. Obviously what is lacking is not just its being given to *the mind* 'in extension'; that is lacking for most sets as well. What is lacking has to do, one might say, with being, and moreover if the underlying intension had the

theory seems to be treated with reserve by many set theorists, although the reason could be just that in deductive power it is inferior to stronger axioms of infinity.

Locutions requiring either classes or satisfaction and truth are frequent in writings on set theory (cf. my, 1974c), but the characteristic informal use is very weak and could be captured by a free variable formalism for classes with very elementary operations on them, as George Boolos' comments on (1974c) reminded me.

[28]The 'straightforward reading' involves replacing 'set' by 'class' at certain points in the standard model-theoretic account of (modal) logical validity, just as in the case of ordinary logic in set theory. The same should be the case for set theories with intuitionistic logic, which are suggested by the same considerations as suggest the modal language. It should not be thought that changing to intuitionistic logic will remove the fundamental dilemmas about quantification over all sets.

fixed, completed existence a proper class lacks then the class would have it as well. However, some general ideas about intensional concepts do have application to this case. If we think of classes as given only by our understanding of the (perhaps indefinitely extendible) language of set theory, then the assumption that impredicative reasoning about classes is valid is rather arbitrary. This way of looking at classes corresponds to thinking of intensions as meanings and of meanings as constructions of the mind. This is the conception that is appropriate to applying intensional logic to propositional attitudes. Alternatively (and here we have a clearer theory) intensions are thought of as individuated by modal conditions, as 'functions from possible worlds to extensions'. This is the conception appropriate to modal logic. It seems neutral with respect to the question of impredicativity.

Let me make a final comment on the *predicative* conception of classes. If we understand a second-order language containing set theory in this way then set existence does for us the work of the axiom of reducibility in Russell's theory of types. For predicates which are high in a ramified hierarchy or which more generally are expressible only by 'logically complex' means, the existence of a set $\{x: Fx\}$ provides a simple equivalent $x \in a$ for a a *name* of $\{x: Fx\}$. Clearly it serves as an equivalent only *extensionally*. In the intensional situations envisaged above the equivalence of $x \in a$ and Fx will not be necessary even if the name a has been introduced by stipulation ('*a priori*' in Kripke's sense; cf. Tharp 1975), and therefore the two predicates will behave differently in intensional contexts. Thus the license for impredicativity given by assuming the existence of sets does not nullify the predicative conception of intensions even for intensions that have sets as extensions. Of course this 'reducibility' does not obtain for predicates that do not have sets as extensions.[29]

In conclusion, I would claim that the above discussion had added something to the explication of the idea that an 'inconsistent multiplicity' is not really an object, since even as an intension it is systematically ambiguous. The task remains to explain whether the ideas of the last two sections are helpful in understanding the 'stages' of the genetic conception and the underlying priority of the elements of a set to the set.

[29]It is commonly claimed that the axiom of reducibility nullifies Russell's ramification of his hierarchy of types. This claim depends on ignoring, presumably on the grounds that nonextensional features of functions are not significant for mathematics, the fact that Russell thought of propositional functions intensionally.

On the other hand it is hard to see what is left of Russell's no-class theory once the axiom of reducibility is admitted. Russell himself says that the axiom of reducibility accomplishes "what common sense effects by the admission of classes" (1908: 167), but he considers the axiom a weaker assumption than the existence of classes. The weakness must consist in the restrictions of the simple theory of types.

V

In the last two sections we sought to avoid using either epistemic concepts or the metaphor of process in trying to understand the conditions for the existence of sets. However, we concentrated largely on the distinction between sets and classes or 'multiplicities' and on discourse about absolutely all sets.

The idea that any available objects can be formed into a set is, I believe, correct, provided that it is expressed abstractly enough, so that 'availability' has neither the force of existence at a particular *time* nor of giveness to the human mind, and formation is not thought of as an action or Husserlian *Akt*. What we need to do is to replace the language of time and activity by the more bloodless language of potentiality and actuality.

Objects that exist together *can* constitute a set. However, we do have to distinguish between 'existing together' and 'constituting a set'. A multiplicity of objects that exist together *can* constitute a set, but it is not necessary that they *do*. Given the elements of a set, it is not necessary that the set exists together with them. If it is possible that there should be objects satisfying some condition, then the realization of this possibility is not as such the realization *also* of the possibility that there be a set of such objects. However, the converse does hold and is expressed by the principle that the existence of a set implies that of all its elements.

The same idea would be expressed in semantic terms by the supposition that we can use quantifiers and predicates in such a way that the range of the quantifiers and the objects satisfying any one of the predicates can constitute single objects, but these objects are not already captured by our discourse. However, this way of putting the matter might be taken to rule out too categorically an 'absolute' use of quantifiers and predicates. Without returning to an ontological characterization such as the Cantorian language of 'existing together', we can say that this is the condition under which quantifiers and predicates obtain definiteness of sense.

Above we suggested that the axiom of power set rests on a sort of principle of plenitude, according to which all the possible subsets of a given set are capable of existing 'at once'. Against what we have just said one might object that there is no intrinsic reason why the 'potentiality' of a set relative to its elements should not be nullified in our theory by a similar principle of plenitude.

The short answer to this objection is that such treatment would lead to contradictions, Russell's paradox in particular. We could apparently consistently assume (as in New Foundations) that the domain of dis-

course is a set in the domain, but then of course there will be other 'multiplicities' of elements of the domain that are not in it.[30]

A further point is that there seems to be an intrinsic ordering of 'relative possibility' in the element-set relation that is lacking for the arbitrary subsets of a given set. A set is an *immediate* possibility given its elements, the sets of which it is an element are at least at another remove. We do of course have conceptions of the 'simultaneous' realization even of infinite hierarchies in this ordering, but such a conception gives the possibility of sets that are still higher.

This observation should remind us that more is involved in the 'iterative conception' of set than the priority of element to set, since in Gödel's words we think of arbitrary sets as obtained by *iteration* of the "operation 'set of'" starting with individuals, and we have not yet dealt with the concept of iteration. To do so adequately would be beyond the scope of this paper. I shall make a few remarks.

First, our strategy has been to use modal concepts in order to save the idea that *any* multiplicity of objects can constitute a set; one makes only the proviso that they 'can exist together', and this proviso I take to be already given by the meaning of the quantifiers unless they are used in a 'systematically ambiguous' way. One saves thereby the universal comprehension axiom as well, though in a form that hardly seems 'naive' any more: In the second-order modal language it would have to be expressed by the statement that for every attribute P there is an attribute Q that is the rigidification of P (note 24 above) and such that

$$\Diamond(\exists y)(\forall x)(x \in y \leftrightarrow Qx). \qquad (4)$$
[31]

However, even with the assumptions needed to obtain a version of the power set axiom we do not obtain greater power than that given by a much more traditional way of saving the comprehension axiom: the simple theory of types. It is clear that without some principle allowing for *transfinite* iteration of something like the above comprehension principle we will not obtain even the possible existence of sets of infinite rank, such as the usual axiom of infinity already requires. For the axiom of infinity, the principle needed is one allowing the conversion of a 'potential' infinity into an 'actual' infinity: we can easily show

[30] In the case of NF, these additional 'multiplicities' would correspond to the proper classes of ML.

If a model of NF is given as a set in the ordinary set-theoretic sense, the domain of the model and the V of the model will of course differ. The membership relation of the model will obviously not be the same as the membership relation 'from outside'.

[31] Thus if in some possible world $(\forall x)(x \in y \leftrightarrow Qx)$ holds, the elements of y are just the objects that *actually* have P. In many cases they will not be just the objects that have P in the world in question.

(5) $$\Box(\forall x)\Diamond(\exists y)(y=x\cup\{x\}),$$

but to use (4) to infer that ω possibly exists, we would need to get from (5) to

$$\Diamond(\forall x)(\exists y)(y=x\cup\{x\});$$

in terms of a set-theoretic semantics for the modal language, the possible worlds containing finite segments of ω need to be collected into a single one.

Second, it is clear that there has to be a priority of earlier to later ordinals, whether this is *sui generis* or derivative from the priority of element to set. One could of course assume a well-ordered structure of individuals, within which there would be no ontological priority of earlier to later elements. The axiom of infinity of *Principia* is such an assumption. To make it is natural enough, unless we assume a relation to the mind is essential to the natural numbers. Then it seems that smaller numbers are prior to larger ones by virtue of the order of time, as in Brouwer's (and apparently also Kant's) theory of intuition.

For reasons indicated above, no such structure can represent all ordinals. In fact larger ordinals seem conceivable to us only by characteristically set-theoretic means such as assuming that there is already a *set* closed under some operation on ordinals.

Third, it seems to me that the evidence of the axiom of foundation is more a matter of our not being able to understand how non-well-founded sets could be possible rather than in a stricter insight that they are *impossible*. We can understand starting with the immediately actual (individuals) and iterating the 'realization' of higher and higher possibilities. It seems that (at least as long as we hold to the priority of element to set) we do not understand how there could be sets that do not arise in this way. Non-well-founded $\in$-structures have been described (simple ones already in Mirimanoff 1917a), but we do not recognize them as structures of *sets* with $\in$ as the real membership relation, even when they satisfy the axioms of set theory.[32] We are at liberty to say that the *meaning* of 'set' is, in effect, 'well-founded set', but that does not exclude the possibility that someone might conceive a structure very like a 'real' $\in$-structure which violated foundation but which might be thought of as a structure of sets in a new sense closely related to the old.

I shall close with a rather speculative comment. The conception of 'inconsistent multiplicities' as indefinite or ambiguous raises a doubt

[32]Perhaps this could be said of trivial variants such as that resulting from identifying individuals with their unit classes. But here of course a slightly modified axiom of foundation holds.

about whether it is appropriate to talk of *the* cumulative hierarchy as most set theorists do. The definiteness of the power set is maintained even though the hope of deciding such questions as the continuum hypothesis and Souslin's hypothesis by means of convincing new axioms has not been realized. However, in this case the idea of the 'maximality' of the power set gives us some intuitive handle on the plausibility of the hypotheses or of 'axioms' such as V=L that *do* decide them.

Maximality conceptions also contribute to the plausibility of large cardinal axioms. Here it seems conceivable in the abstract that we might see the possibility of a cardinal α with a 'structural property' P and of a cardinal β with such a property Q, where these properties are not 'compossible'; that is, we would see (perhaps even in ZF) that such α and β cannot both exist. That would yield two incompatible possibilities of cumulative hierarchies.

This has not happened with any of the types of large cardinals considered in recent years, where it has generally happened that of two such properties one (say P) implies the other, and indeed $P\alpha$ implies the existence of many smaller β such that $Q\beta$. That this is so has seemed rather remarkable; perhaps it is evidence against the views I have advanced.

However, one reason for thinking that 'incompatible large cardinals' will not arise is that by the Skolem-Löwenheim theorem both would reflect into the countable sets. If our confidence in the uniqueness of $\mathfrak{P}(\omega)$ is so great as to lead us to reject the possibility of incompatible large cardinals, one would still wish for some more direct reason for doing so.[33]

[33]I am indebted to Robert Bunn, William Craig, William C. Powell, Hilary Putnam, and Hao Wang for valuable discussions related to this paper. I regret that time did not permit me to follow up Mr. Bunn's remarks on Jourdain's attempt to develop the theory of inconsistent multiplicities.

The concept of set

HAO WANG

1. The (maximum) iterative concept

A set is a collection of previously given objects; the set is determined when it is determined for every given object x whether or not x belongs to it. The objects which belong to the set are its members, and the set is a single object formed by collecting the members together. The members may be objects of any sort: plants, animals, photons, numbers, functions, sets, etc.

According to the iterative concept, a set is something obtainable from some basic objects (such as the empty set, or the integers, or individuals, or some other well-defined urelements) by iterated applications of the rich operation 'set of' which permits the collecting together of any multitude of 'given' objects (in particular, sets) or any part thereof into a set. This process includes transfinite iterations. For example, the multitude of sets obtained by finite iteration is considered to be itself a set.

We understand this concept of set sufficiently well to see, after some deliberation, and in some cases even a great deal of deliberation, that the ordinary axioms of set theory are true for (or with respect to) this concept, and to be able to extend these axioms by proposing additional axioms and recognizing some of them to be true for (or with respect to) it.

The iterative concept involves at least four difficult ideas: the idea of 'given', the idea of collecting together, the idea of 'part' or subset, and the idea of iteration. The idea of iterations implies the potentiality of continuing to any stage (as indexed by a previously given ordinal number)[1] and adds an inductive element to the idea of 'given' (viz. all sets obtained at or before any given stage are viewed as given). The idea of urelement is not difficult for set theory, because we are in this context

Reprinted with the kind permission of the author and publisher from Hao Wang, *From Mathematics to Philosophy,* Routledge and Kegan Paul, 1974, pp. 181–223. Permission for publication in America kindly granted by Humanities Press, Inc.

[1] The reader who is not familiar with the technical concepts of set theory is referred to standard texts such as Kamke 1950 and Fraenkel 1953. For more specialized concepts and results, the reader may consult Hausdorff 1949, Gödel 1940, Bernays 1958, and Cohen 1966.

not interested in what an individual is but rather leave the question open. We do not attempt to determine what the correct urelements are.

It is a basic feature of reality that there are many things. When a multitude of given objects can be collected together, we arrive at a set. For example, there are two tables in this room. We are ready to view them as given both separately and as a unity, and justify this by pointing to them or looking at them or thinking about them either one after the other or simultaneously. Somehow the viewing of certain given objects together suggests a loose link which ties the objects together in our intuition, or a variable object which could be any one of them. In order that our mind may more effortlessly and unwaveringly fix our attention on this variable object, we, it could be suggested, concretize or reify the loosely linked bundle of objects and think of the more determinate range of variability. But then we seem to be forced by the surprising success of the reification to admit that there are certain objective grounds for our ostensively acquired intuition. It may be noted that Cantor discusses briefly the same phenomenon in connection with the move from a potentially infinite to an actually infinite.[2]

We can form a set from a multitude only in case the range of variability of this multitude is in some sense intuitive. This is the criterion for determining whether a multitude forms a set for us. The natural way of getting such intuitive ranges is by the use of intuitive concepts (defining properties). An intuitive concept, unlike an abstract concept such as that of mental illness or that of differentiable manifold, enables us to overview (or look through or run through or collect together), in an *idealized* sense, all the objects in the multitude which make up the extension of the concept, in such a way that there are no surprises as to the objects which fall under the concept. Hence, each intuitive concept determines an intuitive range of variability and therewith a set.

The overviewing of an infinite range of objects presupposes an infinite intuition which is an idealization. Strictly speaking, we can only run

[2] Cantor 1932. All references to Cantor are to this volume of his collected works.

Unterliegt es nämlich keinem Zweifel, dass wir die *veränderlichen* Grössen im Sinne des potentialen Unendlichen nicht missen können, so lässt sich daraus auch die Notwendigkeit des Aktual-Unendlichen folgendermassen beweisen: Damit eine solche veränderliche Grösse in einer mathematischen Betrachtung verwertbar sei, muss strenggenommen das 'Gebiet' ihrer Veränderlichkeit durch eine Definition vorher bekannt sein; dieses 'Gebiet' kann aber nicht selbst wieder etwas Veränderliches sein, da sonst jede feste Unterlage der Betrachtung fehlen würde; also ist dieses 'Gebiet' eine bestimmte aktualunendliche Wertmenge. So setzt jedes potentiale Unendliche, 'soll es streng mathematisch verwendar sein, ein Aktual-Unendliches voraus. Diese 'Gebiete der Veränderlichkeit' sind die eigentlichen Grundlagen der Analysis sowohl wie der Arithmetik und sie verdienen es daher in hohem Grade, selbst zum Gegenstand von Untersuchungen genommen zu werden, wie dies von mir in der 'Mengenlehre' (théorie des ensembles) geschehen ist (1886, pp. 410–11).

through finite ranges (and perhaps ones of rather limited size only). This idealization contains seeds for growth in itself. For example, not only are the infinitely many integers taken as given, but we also take as given the process of selecting integers from this unity of all integers, and therewith all possible ways of leaving integers out in the process. So we get a new intuitive idealization (viz. the set of all sets of integers) and then one goes on.

The concept of all subsets is often thought to be opaque because we envisage all possibilities independently of whether we can specify each in words; for example, just as there are 2^{10} subsets of a set with 10 members, we think of 2^a subsets of a set with a members when a is an infinite cardinal number. In particular, we do not concern ourselves over how a set is defined, e.g. whether by an impredicative definition. This is the sense in which the individual steps of iteration are 'maximum'. It is possible to get other iterative concepts by restricting the operation of going to the next stage, one familiar example being the constructible sets. The (maximum) iterative concept has been discussed by Bernays (1935) [reprinted in this volume] under the name of platonism.

The weakest 'platonistic' assumption introduced by arithmetic is that of the totality of integers... 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... 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. [cf. pp. 259–60, this volume]

What is given at each stage depends on an orderly manner of iteration. Hence, the concept of ordinal number is essential to the iterative notion, in that we use ordinal numbers to index the stages of iteration. Thus, as we generate more and more sets according to the iterative concept, we encounter certain well-ordered sets among the sets generated. The order types of these well-ordered sets determine ordinal numbers which can be used to index (further) stages of the iterations. Given a totality of operations for generating sets, we can also survey all the ordinal numbers obtainable by these operations and introduce new ordinals. In general, for any ordinal number α, given by whatever means, we are permitted to carry the process of iteration to the α-th stage, and regard all the sets generated up to and including the α-th stage as given and proceed further.

The question of urelements involves us in the contrast between sets (or mathematical objects in general) and other objects. Philosophically it is

important to realize that we are faced with a perfectly general situation and that we can begin initially with any collectable multitudes as urelements. There is nothing in the original iterative concept to rule out different kinds of urelements. For example, we can take all physical objects as the urelements, or all elementary particles, or all animals, or all integers, etc. In each case, if we are able to collect the urelements into a set x, we can carry out the process of iteration starting with x (or conceive of a hierarchy of transfinite types with x at the bottom). Since, however, the process of generating further sets from an initial set x of urelements is uniform, with respect to x, i.e. the process remains the same no matter what initial set x we might wish to choose, it is reasonable to consider just one typical general case as we have done in our first explanation of the iterative concept.

Moreover, for the abstract study of sets, it seems convenient to disregard nonsets altogether. This turns out to be feasible, because even if we start from nothing (i.e. neither urelements nor sets) initially, we can get the empty set 0. The use of this artificial special case of an empty set of urelements achieves a convenient purity. As a matter of fact, for the mathematical studies of sets it is customary to require that all members of sets are sets. This restriction excludes sets of tables and elephants, but does not exclude sets of numbers and functions which are identified with certain sets. Under this restriction, we say that a set is a collection of given sets.

On the basis of our explanations of the (maximum) iterative concept of set, we are able to see that the ordinary axioms of set theory (commonly referred to as *ZF* or *ZFC*) are true for the concept.

AE Axiom of extensionality. A set is completely determined by its members; i.e. two different sets may not contain the same members. If x and y have the same members, then $x=y$.

This may be viewed as a defining characteristic of sets (in contrast with properties).

AS Axiom of subset formation (axiom of comprehension). If a multitude A is included in a set x, then A is a set.

Since x is a given set, we can run through all members of x, and, therefore, we can do so with arbitrary omissions. In particular, we can in an idealized sense check against A and delete only those members of x which are not in A. In this way, we obtain an overview of all the objects in A and recognize A as a set.

AP Axiom of power set. All subsets of a set can be collected into a set.

For, if x is given, then all subsets of x are given individually by *AS*. We have, moreover, an intuitive idea of running through with omissions. This general notion, which is on a higher level than its application to each multitude A included in x, provides us with an overview of all cases of *AS* as applied to x. And the overview provides us with the basis of performing the collection to get the power set of x.

In our previous discussion about the urelements, we have reached the conclusion that for the abstract development of set theory we can conveniently disregard the diversity of urelements and, in fact, leave out nonsets altogether, taking an empty set of urelements. Now that we are justified in forming the power set of a given set, we are able to tidy up the iterative process in another direction. The operation of forming the power set of a given set eliminates the need to branch out from a given set x: there are different ways of forming subsets of x; we might otherwise be forced to distinguish between different kinds of subsets of x so that certain subsets of x are collected into one new set, and certain other subsets into another new set, and so on. By using the power set of x, we are able to pull together all subsets of x and summarize the formation of all possible subsets of x in a single new set, viz. the power set of x. In this way, we obtain a standard representation of all single applications of the rich operation 'set of' to any given totality of given objects. There are then no other obstacles against our construing the iterative conception in the sharper form of ranks or types or stages: every set is obtainable at some stage α (an ordinal number) and every stage R_α is obtained from the empty set (of urelements) by iterated applications of the operation 'set of,' which yields all members of the power set of R_β if $\alpha=\beta+1$, and just gathers together all sets obtained at previous stages if α is a limit ordinal number. In other words, if R_α is the totality of sets obtained at all stages before the α-th, then $R_{\alpha+1}$ consists of all the subsets of R_α. For example, $R_0=0$, $R_1=\{0\}$, $R_2=\{0,\{0\}\}$, $R_3=\{0,\{0\},\{\{0\}\},\{0,\{0\}\}\}$, and so on.

The iterative concept implies that we continue the iteration as far as possible; in particular, it implies that, for any given ordinal number α, there is an α-th stage. There is then the problem of getting ordinal numbers to index the stages. For example, we take for granted that we have the finite ordinal numbers to begin with. We are then led, as it happened to Cantor originally, to ω as the limit of all finite ordinal numbers (the natural numbers) and then to $\omega+1$, and so on.

Thus, for each natural number n, we have a stage R_n. But there is no reason to stop there. So we have a further stage R_ω which collects together all the finite stages, as well as stages $R_{\omega+1}$, etc. From the way the stages are obtained, we see that for every set obtained, there is a first

stage at which it appears, and that if there is at least one stage possessing a certain property, then there is a first stage possessing that property.

> *AF Axiom of foundation*. Every set can be got at some stage; or, every nonempty set (or even multitude of sets) has a minimal member, i.e. a member x such that no member of x belongs to the set.

For there is a member x which is got at no later stage than any other member of the set. But all members of x are got at earlier stages and therefore cannot belong to the set.

> *AI Axiom of infinity*. There is an infinite set (for example, R_ω).

> *AC Axiom of choice*. Given any set x of nonempty sets, there is a set which contains exactly one member from each member of x.

Since every member of x is got at an earlier stage than x, all members of members of x are got earlier and any selection from these can be collected together to form a set.

> *AR Axiom of replacement*. If b_x is a set for every member x of a set y, then the union of all these sets b_x is included in a set.

This form of *AR* differs from the more familiar form in two minor aspects: the use of the union and the weakening from being a set to being included in a set. The familiar form is *SAR*: if b is an operation and b_x is a set for every member x of a set y, then all these sets b_x form a set. The differences are introduced for certain esthetic reasons, which are not very relevant to our main interest here. We shall relegate a crude direct justification of *AR*, as well as an explanation of the relation between *AR* and *SAR*, to a footnote.[3] Here, we shall confine our attention to *SAR*.

[3] In very rough terms, we may directly justify the axiom *AR*, on the basis of the iterative concept, in the following manner.

In general, given any set y, we may consider the multitude of all stages $R_{\alpha(x)}$ where x is a member of y and $R_{\alpha(x)}$ is the first stage at which x appears. A reasonable principle for continuing the stages is to permit, for each given set y the collection or merging of all these stages $R_{\alpha(x)}$ into a new set. If, instead of $R_{\alpha(x)}$, we take any given set b_x, it is no less justifiable to collect or merge all the stages where these sets first appear, to get a new set. By this principle, if b_x appears at stage $R_{\beta(x)}$ then the result obtained by merging these stages, for all x belonging to y, contains the union of these sets b_x.

The minor additional procedure of forming the union of a set (i.e. merging the elements of a set) is conceptually a consequence of the intended process of iteration, since all members of members of a set a are given at earlier stages and therefore collected into a set b before the stage at which a emerges. A separate axiom for forming union is often included in formal systems of set theory because the defining properties do not mirror faithfully the intended extensional interpretation. The absorption of this feature into the replacement axiom (as stated above) is meant to render it less conspicuous. An inessential feature of the form of *AR* as stated is to get, instead of the image of y directly, a set which contains it. This amounts to taking, instead of b_x itself, its corresponding stage $R_{\beta(x)}$ so that (the union

Once we adopt the viewpoint that we can in an idealized sense run through all members of a given set, the justification of *SAR* is immediate. That is, if, for each element of the set, we put some other given object there, we are able to run through the resulting multitude as well. In this manner, we are justified in forming new sets by arbitrary replacements. If, however, one does not have this idea of running through all members of a given set, the justification of the replacement axiom is more complex.

Gödel points out that the axiom of replacement does not have the same kind of *immediate* evidence (previous to any closer analysis of the iterative concept of set) which the other axioms have. This is seen from the fact that it was not included in Zermelo's original system of axioms. He suggests that, heuristically, the best way of arriving at it from this standpoint is the following. From the very idea of the iterative concept of set it follows that if an ordinal number α has been obtained, the operation of power set (P) iterated α times leads to a set $P^{\alpha}(0)$. But, for the same reason, it would seem to follow that if, instead of P, one takes some larger jump in the hierarchy of types, e.g. the transition Q from x to $P^{|x|}(x)$ (where $|x|$ is the smallest ordinal of the well-orderings of x), $Q^{\alpha}(0)$ likewise is a set. Now, to assume this for any conceivable jump operation (even for those that are defined by reference to the universe of all sets or by use of the choice operation) is equivalent to the axiom of replacement.[4]

The seven axioms *ESPFICR* will be regarded as making up the ordinary system *ZF* (or *ZFC*) of set theory. The comments above about these axioms are intended to show that we can see them to be true for the iterative concept of set. Somewhat more formally, we can also recapitulate the hierarchy of sets resulting from the iterative concept, by assuming that the ordinal numbers are given initially, as follows.

of) the multitude of all the sets b_x, for x in y, is included in the set of all their corresponding stages $R_{\beta(x)}$. This serves the purpose of avoiding the somewhat inelegant situation of making the more basic axiom of comprehension a consequence of the replacement axiom.

It should perhaps be pointed out that these seven axioms *ESPFICR* are equivalent to other more commonly used sets of axioms taken as making up *ZF*. For a detailed proof, the reader may consult Shoenfield 1967: 240–3. The derivations making up the proof of equivalence go back at least to Zermelo (1930) and Bernays (1958); compare the references under notes 13 and 1.

[4]More explicitly, I would like to add as a supplement, it is a familiar fact that once we have replacement from sets of ordinals to get new sets of ordinals and we permit a stage R_{α} for each given ordinal α, we can get full replacement. And it is easily seen that replacement for sets of ordinals (i.e. given $f(\alpha)$, $\alpha<\beta$, there is γ, $f(\alpha)<\gamma$, for all $\alpha<\beta$) follows from the iteration of jumps (i.e. given f and β, there is γ^{α}, $f(0)<\gamma$, for all $\alpha<\beta$).

Gödel's explanation of the jump operation may also be viewed as a generalization of the way Cantor applies (in the development of his second number class) his second principle of generation according to which, if there is defined any definite succession of ordinal numbers of which there is no greatest, a new number is created which is defined as the next greater to them all (1883, p. 196).

$R_0 =$ the empty set (or, sometimes, the set of integers)
$R_{\alpha+1} =$ the power set of R_{α}, i.e. the set of all subsets of R_{α}
$R_{\lambda} =$ the union of all R_{α}, $\alpha<\lambda$, where λ is a limit ordinal
$V =$ the union of all R_{α}, α any ordinal

In other words, the universe of all sets consists of all x such that x belongs to some R_{α}, α an ordinal. The smallest α such that x belongs to R_{α} is usually called the rank of x. Under this formulation, it is clear that the two difficult ideas are power set and ordinal number. In recent years, much effort has been devoted to finding more ordinals by introducing new cardinals to strengthen axiomatic set theory. In contrast, there has been little progress in efforts to enrich directly power sets (e.g. that of the set of integers) by new axioms. Both endeavors could be viewed as attempts to make our vague intuitive ideas more explicit.

The iterative concept seems close to Cantor's original idea,[5] and has been, in one form or another, developed and emphasized by Mirimanoff (1917a, b), von Neumann (1925), Zermelo (1930), Bernays (1935), and Gödel (1964).

This iterative concept of set is of course quite different from the dichotomy concept which regards each set as obtained by dividing the totality of all things into two categories (viz. those which have the property and those which do not). Following Gödel, one may speak of the two concepts as the mathematical versus the logical. To quote:

> There exists, I believe, a satisfactory foundation of Cantor's set theory in its whole original extent and meaning, namely axiomatics of set theory interpreted in the way sketched below. It might seem at first sight that the set-theoretical paradoxes would doom to failure such an undertaking, but closer examination shows that they cause no trouble at all. They are a very serious problem, not for mathematics, however, but rather for logic and epistemology. [Gödel 1964: 262; 474 in this volume]

Many people have been puzzled by the fact that in an earlier paper on Russell (1944: 215–16 [452 in this volume]), Gödel takes the paradoxes much more seriously. 'By analyzing the paradoxes to which Cantor's set theory had led, he freed them from all mathematical technicalities, thus bringing to light the amazing fact that our logical intuitions (i.e. intuitions concerning such notions as: truth, concept, being, class, etc.) are self-contradictory.' The difference in emphasis, as Gödel explains, is due to a difference in the subject matter, because the whole paper on Russell is concerned with logic rather than mathematics. The full concept of class

[5]Compare the discussions to follow (in particular, notes 6, 8, 9). It may be said that not only the famous 1895 definition in terms of a collection of objects into a whole, but even also the 1883 definition in terms of one and many suggest strongly the iterative concept.

(truth, concept, being, etc.) is not used in mathematics, and the iterative concept, which is sufficient for mathematics, may or may not be the full concept of class. Therefore, the difficulties in these logical concepts do not contradict the fact that we have a satisfactory mathematical foundation of mathematics in terms of the iterative concept of set. In relation to logic as opposed to mathematics, Gödel believes that the unsolved difficulties are mainly in connection with the intensional paradoxes (such as the concept of not applying to itself) rather than with either the extensional or the semantic paradoxes. In terms of the contrast between bankruptcy and misunderstanding as considered below, Gödel's view is that the paradoxes in mathematics, which he identifies with set theory, are due to a misunderstanding, while logic, as far as its true principles are concerned, is bankrupt on account of the intensional paradoxes.[6]

One feels vaguely that the iterative concept corresponds pretty well to Cantor's 1895 'genetic' definition of set:[7] 'By a "set" we shall understand any collection into a whole *M* of definite, distinct objects *m* (which will be called the "elements" of *M*) of our intuition or our thought.' We are naturally curious to know a little more about the development of Cantor's concept and its relation to the iterative concept.

In 1882, Cantor explains that a set of elements is *well defined,* if by its definition and by the logical principle of excluded middle we must recognize as internally determined whether any object of the right kind belongs to the set or not.[8] One is inclined to think that the concept of set implicit in this context is closer to the logical concept rather than the mathematical one. In the next year, a set is defined, with references to Plato's notion of ideas and other related concepts, as[9] 'every Many, which can

[6]In order to prevent any misinterpretation of this remark, Professor Gödel suggests adding the following. 'This observation by no means intends to deny the fact that *some* of the principles of logic have been *formulated* quite satisfactorily, in particular all those that are used in the application of logic to the sciences including mathematics as it has just been defined.'

[7]'Unter einer "Menge" verstehen wir jede Zusammenfassung M von bestimmten wohlunterschieden Objekten *m* unserer anschauung oder unseres Denkens (welche die "Elemente" von M gennanten werden) zu einem Ganzen' (1895, p. 282).

[8]Eine Mannigfaltigkeit (ein Inbegriff, eine Menge) von Elementen, die "irgendwelcher Begriffssphäre angehören, nenne ich *wohldefiniert,* wenn auf Grund ihrer Definition und infolge des logischen Prinzips vom ausgeschlossenen Dritten es als *intern bestimmt* angesehen werden muss, *sowohl* ob irgendein derselben Begriffssphäre angehöriges Objekt zu der gedachten Mannigfaltigkeit als Element gehört oder nicht, *wie auch,* ob zwei zur Menge gehörige Objekte, trotz formaler Unterschiede in der Art des Gegebenseins einander gleich sind oder nicht. Im allgemeinen werden die betreffended Entscheidungen nicht mit den zu Gebote stehenden Methoden oder Fähigkeiten in Wirklichkeit sicher und genau ausführbar sein; darauf kommt es aber hier durchaus nicht an, sondern *allein* auf die *interne Determination,* welche in konkreten Fällen, wo es die Zwecke fordern, durch Vervollkommnung der Hilfsmittel zu einer *aktuellen (externen) Determination* auszubilden ist" (p. 150).

[9]'Unter einer "Mannigfaltigkeit" oder "Menge" verstehe ich nämlich allgemein jedes

be thought of as One, i.e. every totality of elements that can be united into a whole by a law.'

According to Fraenkel, Cantor had discovered the so-called Burali-Forti paradox no later than 1895, i.e. at least two years before Burali-Forti's publication, and had communicated it to, among others, Hilbert in 1896 (see p. 470 of Cantor's *Works*). This discovery may have something to do with the 'genetic' element in the famous 1895 definition. According to Zermelo (in Cantor 1932, p. 352, footnote 9), part of the reason why Cantor, in his treatise of 1895–7, deals extensively with the second number class rather than with all cardinal numbers was Cantor's awareness of the 'Burali-Forti paradox.' This may also explain why Cantor, in his 1895 paper, spoke of desiring to show that all cardinals form a well-ordered set 'in an extended sense' (p. 295). A concrete proposal along the line of distinguishing sets and classes was made in Cantor's letter to Dedekind in 1899 (pp. 443–4), not published until 1932.

There are also other differences between Cantor's outlook and the current one. But these seem to belong more appropriately to a footnote.[10]

Viele, welches sich als Eines denken lässt, d. h. jeden Inbegriff bestimmter Elemente, welcher durch ein Gesetz zu einem Ganzen verbunden werden kann, ...' (p. 204). The parenthetical explanations of sets in the contexts of defining cardinality in 1884 (p. 387) and 1887 (p. 411) do not seem to add anything.

[10]Cantor does consider point sets (sets of real numbers) and sets of integers, as well as functions of point sets. But in his general development one sees more a theory of transfinite numbers than a set theory. He quickly extracts cardinal and ordinal numbers from sets and devotes most of his attention to these infinite numbers. The impression is that he believes there is a great variety of objects so that no neat structure can be imposed on all sets above and beyond that imposed by such basic notions as cardinal and ordinal numbers.

For Cantor, cardinals and ordinals are not sets but general concepts or universals abstracted from sets of equal cardinality and isomorphic well-ordered sets. For example, the cardinality of a set *x* is what is common to all sets 'equivalent' to *x* (p. 141, 1879; p. 387, 1887; p. 283, 1895; p. 444, 1899). Two sets are equivalent if there is a one-one correspondence between them. Cantor does work freely with numbers as objects and forms sets of them. For example, the first infinite cardinal is that of the set of finite cardinals.

On the other hand, we certainly cannot identify a number with its extension and take it as a set in our universe of sets. For example, the extension of the universal 1 is as large as the universe of all objects (including all sets) since, for each x, $\{x\}$ is a set of cardinality 1. Since the universe of objects consists of sets and urelements and numbers are objects but not sets, Cantor seems to treat them as urelements. This creates some problem with regard to the iterative concept of set. How do we assign ranks to the numbers? A natural suggestion is to give all urelements the rank 0 and the rank of a set would as in the current form, be determined inductively by the membership relation. But then R_0 which is a set would be too large and contain what Cantor calls an inconsistent manifold. One alternative would be distributing numbers into different ranks.

There is indeed a natural way of doing this, as has been carried out by Mirimanoff (1917a, b), Zermelo (unpublished work of 1915, see Bernays 1941: 6), and von Neumann (1923). Each ordinal number is identified with a canonical set representing it: take the empty set as 0, $\alpha\cup\{\alpha\}$ as $\alpha+1$, the limit ordinal as the set of representatives of all preceding ordinals. Cantor did not make this convenient identification. But it seems likely that Cantor thinks of an open domain of objects which are not sets, such as physical objects, With regard to the task of setting up the axioms of set theory (including the search for new axioms), we can distinguish two questions, viz. (1) what, roughly speaking, the principles are by which we introduce the axioms, (2) what their precise meaning is and why we accept such principles. The second question is incomparably more difficult. It is my impression that Gödel proposes to answer it by phenomenological investigations.

In connection with the first question, Gödel suggests the following summary of the principles which have actually been used for setting up axioms. It is understood that the same axiom can be justified by different principles which are nevertheless distinct in that they are based on different ideas; for example, inaccessible numbers are justified by either (2) or (3) below. The five principles to follow are illustrated by the discussion so far and the section below on new axioms and criteria of acceptability.

(1) Existence of sets representing intuitive ranges of variability, i.e. multitudes which, in some sense, can be 'overviewed' (see above).

(2) Closure principle: if the universe of sets is closed with respect to certain operations there exists a set which likewise is. This implies, e.g., the existence of inaccessible cardinals and of inaccessible cardinals equal to their index as inaccessible cardinals.

(3) Reflection principle: the universe of all sets is structurally undefinable. One possibility of making this statement precise is the following: The universe of sets cannot be uniquely characterized (i.e. distinguished from all its initial segments) by any internal structural property of the $\in$-relation in it, expressible in any logic of finite or transfinite type, including infinitary logics of any cardinal number. This principle may be considered as a generalization of (2). Further generalizations and other precisations are in the making in recent literature.

(4) Extensionalization: axioms such as comprehension and replacement are first formulated in terms of defining properties or relations. They are extensionalized as applying to arbitrary collections or extensional correlations. For example, we get the inaccessible

experiences, universals, properties, and whatnot. Hence, he would not have examined the idea that they can all either be put into R_0 or be given other appropriate ranks in a natural way. However that may be, Cantor's development of set theory as a mathematical subject can be embodied in a framework of pure sets, once the identification of numbers with suitable sets is made.

It is perhaps natural to think of the urelements as forming a set. If, on the other hand, one assumes that there are as many sets (or numbers) as urelements, one might perhaps modify the iterative model by defining $R_\alpha(a)$ relative to each set a of urelements such that $R_0(a)=a$ but $R_{\alpha+1}(a)$ and $R_\lambda(a)$ are defined as above.

numbers by (2) above only if we construe the axiom of replacement extensionally.

(5) Uniformity of the universe of sets (analogous to the uniformity of nature): the universe of sets does not change its character substantially as one goes over from smaller to larger sets or cardinals, i.e., the same or analogous states of affairs reappear again and again (perhaps in more complicated versions). In some cases it may be difficult to see what the analogous situations or properties are. But in cases of simple and, in some sense, 'meaningful' properties it is pretty clear that there is no analog except the property itself. This principle, e.g., makes the existence of strongly compact cardinals very plausible, due to the fact that there should exist generalizations of Stone's representation theorem for ordinary Boolean algebras to Boolean algebras with infinite sums and products.

2. Bankruptcy (contradiction) or misunderstanding (error)?

The reactions of Frege and Cantor to the paradoxes were sharply different and can be described as the bankruptcy theory versus the misunderstanding theory. The difference can undoubtedly be attributed completely to their different conceptions of set (the logical versus the mathematical notion). A related reason may perhaps be described as the difference between viewing sets from outside (Frege) and actually doing set theory (Cantor). Typically in philosophical discussions on the foundations of a subject, the emphasis of insiders and outsiders tends to differ. Even when the same statements are endorsed, quite different things could be intended. The meaning of methodological statements can be so indefinite that it is sometimes not easy to reconcile what a specialist says with what he does.

For example, Cantor, Zermelo, Mirimanoff, and von Neumann all seem to have basically the same conception of set, at least with regard to properties of sets which are implicit in the familiar axioms of today. Yet what they say sounds quite different. Cantor apparently thinks that the paradoxes are paradoxical only because the concept of set is not correctly understood (see, e.g., p. 470, letter of 1907). Zermelo construes the paradoxes as necessitating some restrictions on Cantor's 1895 definition of set:

It has not, however, been successfully replaced by one that is just as simple and does not give rise to such reservations. Under these circumstances there is at this point nothing left for us to do but to proceed in the opposite direction and, starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline. In solving the problem we must, on the one hand, restrict these principles sufficiently to exclude all contradictions and, on the other, take them sufficiently wide to retain all that is valuable in this theory. (1908: 261)

According to Mirimanoff,

One believes and it appears evident, that the existence of individuals must imply the existence of sets of them; but Burali-Forti and Russell have shown by different examples that a set of individuals need not exist, even though the individuals exist. As we cannot accept this new fact, we are obliged to conclude that the proposition which appears evident to us and which we believe to be always true is inexact, or rather that it is only true under certain conditions. (1917a: 38)

In discussing attempts to axiomatize set theory, von Neumann (1925: 219–40 and 1961–3, vol. 1) emphasizes an arbitrary element:

Naturally, it can never be shown in this way that the antinomies are actually excluded; and much arbitrariness always attaches to the axioms. (There is, to be sure, a measure of justification of these axioms in that they turn into evident propositions of naive set theory, when the axiomatically meaningless word 'set' is taken in Cantor's sense. But what is deleted from naive set theory – and to avoid the antinomies it is essential to make some deletion – is absolutely arbitrary.) (1961–3, 1: 37)

In the extreme cases, the proponents of the misunderstanding theory propose to uncover flaws in seemingly correct arguments, while the bankruptcy theorists find our basic intuition proven to be contradictory and seek to reconstruct or salvage what they can, by ad hoc devices if necessary. The basic intuitive concept is often called naive set theory and identified with the belief in an absolute comprehension principle according to which any property defines a set. That some notion like this was actually seriously developed by Frege was a historical accident often advanced as evidence that we do have such a contradictory intuition. The principle, if correct, in fact appears to be the sort of thing which belongs to the domain of logic. Hence, it is much easier to understand Frege's enthusiasm over the thesis of reducibility of mathematics than that of his followers. Viewed in the light of Cantor's development of set theory, however, it is not at all clear that we do have such a contradictory intuition. It seems more appropriate to say that we have an inexact intuition which leads to the iterative conception as we notice the paradoxes and the flaws in them. It is, therefore, debatable whether we have such an intuition to begin with. But perhaps this could easily degenerate into a terminological debate.

Now, a stronger assertion is that the inconsistent concept is the only intuition of set we have. Hence, once the concept is seen to be wrong, we are left with nothing but the task of reconstruction as described in the above quotation from Zermelo. Taken literally, Zermelo's two constraints of not too narrow and not too broad are rather weak and leave room for much arbitrariness in that many mutually incompatible set theories are possible solutions. Moreover, there is implicitly an additional arbitrariness in deciding what results are to be taken as data for the reconstruction since the notion of 'valuable' (perhaps also implying 'reliable') is clearly ambiguous. Assuming that we have a good idea what the data are, the task as described sounds much like a combinatorial puzzle which in principle admits of diverse solutions. Even in the empirical sciences, such a situation does not satisfy the intellect. For example, the eightfold way theory of elementary particles has such a flavor but most people look for either a refutation of the theory or more basic principles from which the theory can be deduced. It will be said that, if in fact we do not have good intuitions about sets, then our wish for a stable solution of the paradoxes is futile. And it is not hard to see why the 'tough-minded' position of crying bankruptcy has a certain appeal: it gives the impression of greater 'clarity,' a defiance of tradition, and independence from the slippery matter of intuition.

But the fact is we dò arrive at a fairly stable iterative concept of set, whether or not we agree that this is the only original intuitive notion of set to begin with. And this concept was also implicit in the works of Zermelo and von Neumann who in the quoted contexts speak as bankruptcists, but used their good intuitions about sets in setting up their axiom systems. In any event, even if we agree that our intuition did once lead to contradictions, that fact does not justify the view that we run a high risk of self-contradiction whenever we use our intuition. The striking fact is that people do set theory by extensive appeals to their intuition and there is a practically universal agreement on the correctness or incorrectness of the results thus obtained, as results about sets. The iterative concept of set is an *intuitive* concept and *this* intuitive concept has led to no contradictions.

This is not to say that we have made the iterative concept fully exact and explicit: there remain problems about the indefiniteness of the concept of definite property and one-one correspondence, the range of ordinals we can envisage, the limitations of the axiomatic method. It is not even denied that, within this framework, there is room to experiment with new axioms and be open-minded as to the choice between alternatives. But the historical and conceptual matters sketched so far seem to discredit the bankruptcy view according to which, even today, the fundamental problem of the foundations of set theory remains the solution of the antinomies.

There is a related distinction between formalists and realists (or objectivists). As these positions get further refined, there is a certain convergence of views on matters regarding the correctness of results, even though there is a difference in choosing different problems to work on, for example, a preference by formalists for constructible sets and relative consistency results over speculations on and derivations from very large cardinals. Another difference is in the matter of working habits, so that one might be an avowed realist but think mainly formalistically, while an avowed formalist may use intuitions very efficiently in doing set theory and yet claim that set theory has only a formalistic model. In any case, any serious formalistic position does accept that we have perfectly reliable intuitions with regard to integers and some would claim intuitionistic reasoning as mostly evident. In other words, a formalist position on sets is given more content by contrasting set theory with other (usually more restricted) areas which do have more than a formal subject matter. One also thinks of degrees of reliability. The objectivistic position is a modification of realism with the goal of avoiding a number of extraneous difficulties with mathematical objects.

3. Objectivism and formalism in set theory

Different philosophical positions may be reached either by using the same data or by using different data. With the same data, the disagreement may often be apparent rather than real. More (relevant) data ought to be an advantage. It is not always easy to determine what is acceptable as data. For example, working informally with *ZF* is easier than with *NF*;[11] with *ZF*, while one pursues the argument without regard to formalization, the end results usually come out all right and can, if one wishes, be made into formal proofs from the axioms. This is at least in part due to our ability to think in terms of the intuitive models rather than the formal axioms. One might expect that the finite axiomatization of *NF* would yield fairly directly a contradiction, but the enumeration of all objects turns out to use an unstratified formula. Also, the ability to work with, e.g., a set which is a standard model of *ZF*, an assumption not formally provable in *ZF*, not only yields correct results but facilitates the flow of our arguments. Another point is the convergence of theories which at one stage were regarded as based on fundamentally different ideas.

[11] The system of Quine 1937: 70–80.

For this reason, it cannot be said that our belief in the superiority of one set theory is merely a result of sociological factors such as familiarity, conformity, and respect for authority. It seems to be unquestionable that we have come to accept axioms of extensionality, replacement, choice, and foundation. Perhaps more doubtful is the acceptance that the hypothesis of constructibility is false and that we suspend judgment on the truth or falsity of the hypothesis of measurable cardinals. Sometimes we accept an axiom once stated, sometimes it takes a fairly long time before an axiom is accepted (e.g. the axiom of choice), but at the end we reach an agreement. We do not have two camps of comparable force such that one accepts an axiom, the other rejects it. The agreement also persists in time, i.e. the community does not oscillate from one day to the next. There is also a pretty good agreement on when to suspend judgment. By empirical induction we expect similar agreement in the future, and similar persistence.

It is often not easy to give precise reasons why a certain axiom is accepted or rejected. And also, what is accepted need not always be reflected faithfully in a formal (statement of an) axiom. Hence, the possibility of refinement and modification is not excluded. And the search for more articulate explanations of these empirical facts is of philosophical interest. But it does not follow that unless clear reasons can be given, these surprising phenomena of agreement and coherence must be considered illusions. Here we tread on a thin line between passive acceptance of the fashion and capricious irreverence.

It appears that if mathematics deals with objects, then every mathematical proposition is true or false. There is a natural tendency to think of objects and models. On the other hand, we may wish to say that what is more basic is the successor or the membership relation and that they have certain properties. This does not confine us to any fixed formal systems. In the first place, the rule of induction, for example, is usually taken in an informal way or, in other words (what is really the same thing), taken as a second order statement. As is well known, we then have again the standard model. This is probably one way of upholding objectivism without relying on objects. In the second place, it is not implied that we know all the properties in advance. It is not excluded that we may in the process of studying the subject further come upon and accept new axioms. Perhaps this does leave room for the possibility that there is some yet undiscovered limitation which will show that, for example, the continuum hypothesis is undecidable in a certain stronger sense. The limiting case would be that there are certain absolutely undecidable propositions in set theory. But nobody knows how to work with the concept of absolute undecidability.

A very different position would be: since the continuum hypothesis is undecidable in *ZF, therefore,* the question of its truth loses meaning. In other words, axioms and theorems are true, but undecidable propositions can neither be true nor be false. Let us refer to this mixed position as *M*. A more radical and perhaps more consistent position says that it makes no sense to speak of propositions of set theory (or, according to another extreme viewpoint, any mathematical propositions) as having a truth value (or that they are not really propositions), axioms and hypotheses being in the same boat. This radical thesis depends either on a recommendation to use the word 'true' in a special way or on the contention that we have no reasonable intuitive concept of set at all. We shall not delay over it but confine our attention to the less radical mixed position *M*.

It is not easy to understand the position *M* in any coherent way. The axioms can only be true on account of an interpretation of the concepts involved. In order that the interpretation withhold judgment on undecidable propositions, the axioms would have to capture fully the 'interpretation.' This means, among other things, that we must not confine ourselves to interpretations in the ordinary sense of two-valued models because in such models every proposition is either true or false. Of course, with the usual axioms of number theory and set theory, we do believe that they do not capture completely our intended interpretations of the central concepts.

The historical origin of this curious position *M* is somewhat complicated. The desire to avoid occult qualities and operate with concrete material as much as possible leads to a delight in formal systems as syntactical objects. Perhaps a transfer of the dubious verifiability theory of meaning is made so that verifiability and falsifiability of propositions of set theory are identified with provability and refutability in a formal system. Apart from other difficulties with general verifiability theory, this viewpoint has its own problems: the limitations of formalization, and the unexplained source of the intrinsic meaningfulness of the axioms and theorems of a formal system.

A different line of defending this mixed position is to argue that propositions of set theory have no independent meaning but only derive their meaning from a superstructure, perhaps useful as a summary or for the economy of thought, which is not based on direct intuitions but on how efficiently one can get back ordinary mathematics from it. From this point of view, there is quite a bit of arbitrariness in our choice of formal systems of set theory, and if a system is adequate for ordinary mathematics, the meaning of the axioms is derived from its consequences in more meaningful areas as a sort of gift. Theorems of set theory which are derivable in the formal system and stay in the superstructure get in turn

their meaning from the axioms. Hence, undecidable propositions of set theory cannot have meaning since the only possible source of meaning for them (viz. provability or refutability) is barred.

This viewpoint cannot account for the relatively stable iterative concept of set and makes the question of truth of the propositions of set theory quite thoroughly a relative matter, depending on which formal system one chooses to use. For example, the axiom of choice, the nonexistence of a universal set or a complement set of every set, the existence of any subset of ω definable only by nonstratified formulas are each true in *ZF* but false in *NF*. Also, the uninhibited comparative study of different systems becomes somewhat of a mystery unless we have an intuitive set theory which we can use with no conscious regard toward formalization.

A favorite example against the pragmatic view that we accept an axiom because of its elegance (simplicity) and power (usefulness) is the constructibility hypothesis. It should be accepted according to the pragmatic view but is not generally accepted as true. Indeed, it is likely to be false according to the iterative concept of set. Basically, it is felt that the pragmatic view leaves out the criterion of intuitive plausibility. The constructibility hypothesis is not plausible in itself and, moreover, many of its consequences are not plausible. For example, it implies the existence of a definable well-ordering of the real numbers and fairly simple uncountable sets without perfect subsets; and these consequences are dubious – they have been said to be contrary to the intuition of ordinary mathematics. It implies a strange pattern of reduction theorems with regard to projective sets.[12] It is not a conceptually pure proposition because it allows ordinal numbers definable only by impredicative definitions or not definable at all, but proceeds to reject all further uses of impredicative definitions. The central argument is, perhaps, that by intention we view sets as arbitrary multiplicities regardless of how or if they can be defined. Hence, it is extremely unlikely that constructible sets, which are essentially the ordinal numbers only, give us all arbitrary sets.

Given this initial implausibility, one may be inclined to view with favor certain propositions which contradict the constructibility hypothesis. In particular, the existence of measurable cardinals is one such proposition, and it implies that there are only countably many constructible sets of integers. On account of the prior belief that the constructibility hypothesis is highly restrictive, this conclusion is seen as further evidence that it is false and as evidence that the measurable cardinal hypothesis has plausibility.

[12]Compare, e.g., D. A. Martin 1968: 687–9.

It has been suggested that possibly all sets are ordinal definable because we may have so many ordinal numbers that the collection turns out to be sufficiently rich. When this argument is applied to defend the constructibility hypothesis, we have a further difficulty in that higher ordinals give no more lower sets. For example, all constructible sets of integers are obtained at stage ω_1 and no large cardinals will change the situation. Here, one would perhaps wish to say that there are a lot of countable ordinals. If there are actually enough countable ordinals to make it true that all sets of integers are constructible, then, of course, the continuum hypothesis would also be true. On the other hand, there are familiar ways of foiling the constructibility axiom while retaining the continuum hypothesis. This is entirely in line with our belief, further substantiated by, though probably not completely dependent upon, the truth of the continuum hypothesis for constructible sets, that we are, relative to our present knowledge, more ready to deny the constructibility hypothesis than the continuum hypothesis.

There are different alternatives to the strong proposition that undecidability in *ZF* implies meaninglessness, or the related proposition that the *ZF* axioms constitute an 'implicit definition' of the concept of set. One alternative is to say that we can never know enough to conclude definitely whether the continuum hypothesis (or the constructibility hypothesis) is true or false. This position would permit our extension of *ZF* to include inaccessible and Mahlo numbers but exclude the possibility of finding clear axioms to decide the continuum hypothesis. A somewhat different alternative would be to say that at least ideas which we have today such as large cardinals cannot possibly lead to a decision on the continuum hypothesis. The position hardest to refute is perhaps that we have simply to withhold judgment: admittedly, as time goes on, we can discover new facts about sets, we can decide more propositions; but, for all we know, we may never be able to decide the continuum hypothesis. One feels uncomfortable if this is put forward as an empirical prediction. Otherwise we would like to see some general arguments. For example, taking into account the diverse possible ways in which languages can grow, we may feel that there are potentially uncountably many questions we can ask about sets. But, certainly we cannot answer uncountably many questions. Hence, why should the continuum hypothesis not be among the unanswerable ones? This can be answered by pointing out that even if we cannot answer all questions, we may be able to answer any question which is singled out as an object of special attention.

Since the continuum problem is to determine the number of sets of integers, it seems reasonable to expect that, barring surprising coincidence, we can only settle the question after we have determined what

objects are to be numbered (what sets of integers are allowed) and on the basis of what one-one correspondences (compare Gödel 1964: 266) [478 in this volume]. But then we seem to be in a difficulty since thus far the determinations we can specify by precise axioms would tend to contradict the intended arbitrary character of sets and one-one correspondences. For example, this is the situation with the notion of constructible sets: we do not regard the continuum hypothesis as shown to be true because it follows from the constructibility hypothesis.

The general limitations of language and formal systems might also suggest that no plausible axiomatization of set theory is likely to be sufficiently refined to determine the exact size of the continuum. It may be the case, for example, that no plausible axioms of set theory will yield the continuum hypothesis, or determine a different specific cardinality for the continuum. But since we have no way of surveying all correct axioms of set theory, there is little likelihood that such a proposition will be established directly rather than approximated by scattered negative results, such as $2^{\aleph_0} \neq \aleph_1$, $2^{\aleph_0} \neq \aleph_4$, etc.

Reasons for believing the continuum hypothesis to be false have been put forward, and are regarded widely as unclear. If one believes the negation of the continuum hypothesis, then, of course, no formal system which includes only true axioms and is consistent with the continuum hypothesis can decide it. This had been used by Gödel as a reason for believing the continuum hypothesis undecidable in *ZF*, before P. J. Cohen established the fact.[13] According to Gödel (1964: 267) [479 in this volume], the continuum hypothesis has implausible consequences. For example, there are results which give uncountable sets which intuitively seem to contain very few members or are highly scattered (e.g. uncountable sets which are meager on every perfect set). But the continuum hypothesis implies that these sets are of the same size as the continuum. The uneasiness about such evidence is based on the feeling that most people do not have a well-developed intuition of large and small with regard to infinite sets apart from the actual development of set theory. On the other hand, it cannot be excluded that someone might have such intimate knowledge so that, for example, he can separate out the errors coming from using the preset-theoretical intuitive concept of largeness. With regard to the matter of intuition, Gödel notes a current fashion against the appeal to intuition and a consequent lack of practice in the conscious use of intuitions. He points out that intuition does not at all mean what first comes to mind but can and should be cultivated.

Some set theorist states that if $2^{\omega} = \omega_1$, then there must be a surprisingly

[13]See Gödel 1947 and Cohen 1966, 1963a: 1143-8, and 1964: 105-10.

delicate balance between the reals and the countable ordinals. But such a remark would be more forceful if it were used against $2^{\omega} = \omega_{17}$ say. As it is, one might say that, for all we know, 2^{ω} might be ω_2, or the first inaccessible number, or real-valued measurable, and that ω_1 is, for all we know, about as reasonable a candidate as any of these.

We do not argue for any strong sharp conclusions but rather try to apply what might be called the dialogue method to determine the limitations of one-sided views. For example, we are not able to establish in any clear sense the thesis that the set-theoretical concepts and theorems describe some well-determined reality, in which Cantor's conjecture must be either true or false. Yet the somewhat indeterminate meanings of the primitive terms of set theory as explained in the iterative notion are accepted as sound. According to Gödel (1964: 272) [484–5 in this volume]: 'The mere psychological fact of the existence of an intuition which is sufficiently clear to produce the axioms of set theory and an open series of extensions of them suffices to give meaning to the question of the truth or falsity of propositions like Cantor's continuum hypothesis.' By the phrase 'give meaning to the question,' Gödel means that there is a good chance of finding a unique answer to the question which will be accepted by all or most of those who are acquainted with the question.

The attraction of the dialogue method is perhaps due at least in part to the fact that what is most interesting in philosophy is not general conclusions but the meaning and limitation of these. Hence, we arrive at the content of these general statements by dialogues. For example, the fact that we have no inconsistency may be due to our limited range of activity relative to all formally possible proofs, and, in addition, our tendency to give proofs which can be interpreted in different frameworks.

It is more natural, certainly for most mathematicians, to deal with objects and models rather than formulas and formal systems. One might wish to claim this is just a shorthand way of doing things even though set theory is a formal game based on analogy and hasty generalizations. The central weakness of this position is of course its apparent inability to explain how the purely formal set theory can hang together so well.

It has been claimed that we have an informal consistency proof of set theory based on considerations about formulas. The point can be illustrated by thinking about the second order arithmetic. Let us try to find directly a countable model for the formal system. We do not have to worry about the fixed sets defined by conditions involving only integers. Now we consider the countably many formulas which contain variables over sets. For each statement $m \in x$ (i.e. $m \in \hat{m}\phi_n m$), we may attempt to try out the two possibilities of being true and being false, adding more set terms to satisfy the impredicatives in ϕ_n or $\neg\phi_n$. In this way, we would

arrive at an intricate graph tree with countably many nodes. For each numeral m and each set term t, we have a formula $m \in t$ (a node) and two branches according as it is taken to be true or false. The truth and falsity of these countably many atomic formulas interact in a complicated way. The problem of consistency is to have a consistent selection of truth values for these atomic formulas, i.e. each formula gets a unique truth value so that all the defining conditions are satisfied.

Viewed in this way, we do not seem to have any good intuition that there must be such a model. In fact, we would find it very surprising if the combinatorial facts resulting from such a formalist outlook on the axioms come out right. In any case, it seems unreasonable to use such a picture with such apparently uncertain outcomes as a means of defending the formalist position.

Alternatively we may follow Gentzen and attempt to prove by transfinite induction that no proof can give a contradiction. In general, whether a formula $n \in t$ is true depends on whether certain other atomic formulas are true or false. If we could obtain an ordering of the degrees of impredicativities, we would be able to get the induction going. But the circular element in the impredicative definitions seems to suggest that we can only get such an ordering in some artificial way, perhaps by assuming what we wish to prove. In fact, it seems that this type of consideration, rather than increasing our belief in the formalist position, has the tendency of suggesting that the platonic picture is the only foothold, vague as it is, we can fall back on.

4. New axioms and criteria of acceptability

Consider first the conditions for accepting a hypothesis in set theory (axiom of choice, hypothesis of measurable cardinals, or some suitably restricted hypothesis of determinacy) as true. Two basic criteria are intrinsic necessity and pragmatic success. The former is related to but perhaps sharper than intuitive plausibility. The latter has various ramifications. One condition is to produce correct lower-order consequences, known (confirmation) and unknown (prediction), for example, about sets of reals, reals, and integers. Another condition is to supply powerful methods of solving problems and even methods which unify diverse results and go beyond them. It is also desirable that the hypothesis be easy to state and to understand. Briefly, we may speak of confirmation, prediction, power, unification (and therefore 'explanation'), and simplicity. The elements of power and unification also contain the component of elegance. Of course, these conditions are neither necessary nor sufficient, since the complicated notion of intrinsic necessity has to dominate and since we may be willing to accept as true hypotheses satisfying only some of these conditions. Of course, there is nothing like a set of quantitative measures by which we can calculate how well each hypothesis fares according to these criteria.

One might wish to view pragmatic success as merely an intermediate criterion for screening candidates for new axioms and require that these candidates eventually pass the test of intrinsic necessity. But there is then the question how invariant the notion of intrinsic necessity is. The iterative concept is admittedly not perfectly clear. Some new axioms may be seen as rendering more exact what we intend, while others may extend or modify our notion in some natural way. If intrinsic necessity is the only way to qualify a hypothesis as an axiom, then there would seem to be no need of comparing set theory with physics. But if pragmatic success can also make a proposition true, one might wonder whether the meaning of the word 'true' is not stretched. To reply to this, we seem to be in a position to say that if pragmatic success is sufficient to make physical hypotheses true, why not also hypotheses in set theory?

There appears to be a sharp contrast between arbitrarily choosing to call a hypothesis true and axioms being forced upon us. Looking backwards, it seems fair to say that axioms which we have accepted so far were forced on us. Therefore, it seems reasonable to expect that we will only accept a hypothesis as true in the future, if the evidence forces it upon us. We do not have a clear idea how such forcing will take place. Moreover, our axioms now all seem to be justified on intrinsic necessity alone. This again suggests that we might choose to wait till existing and future hypotheses achieve the state of intrinsic necessity relative to our understanding before accepting any of them as axioms.

The same data can be interpreted in opposite ways. It has been claimed that the axiom of choice was elevated to the status of an axiom only because of repeated exposure and the psychological reluctance to tolerate central undecidable propositions. One would then have to say that the conceptual justification is no more than an ad hoc rationalization. In the same vein, objects may be regarded as convenient metaphors for discussing formulas and properties.

In any case, it does not seem reasonable to call our choice 'arbitrary,' since we feel we have good reasons for making certain choices and there is a surprising degree of agreement among people who have thought about alternative hypotheses. The two serious positions would seem to be: (1) by accepting a new axiom, we change or extend our concept of set (change the meaning of the word 'set'), and the meaning and truth of the new axiom is determined by the changed concept; (2) we have all along the same concept of set and we accept the new axiom because we have

discovered new facts about sets. There is a strong temptation to say that there is no genuine disagreement, because there is no sharp distinction between changing knowledge by changing meaning and changing knowledge by acquiring new information.

Nobody denies that our intuition of set develops. One should like to be open-minded and allow for the possibility of revoking or modifying our axioms if, for example, contradictions arise or the content of some axiom is rendered more precise by new findings. We cannot, however, disregard our experience so far which seems to indicate a measure of stability leaving little room for difficult choices. We may or may not be successful in explaining satisfactorily the apparently surprising degree of coherence and agreement in set theory, but it is important that we do not belittle this fact on the basis of preconceived philosophical ideas.

Intrinsic necessity depends on the concept of iterative model. In a general way, hypotheses which purport to enrich the content of power sets (say that of integers) or to introduce more ordinals conform to the intuitive model. We believe that the collection of all ordinals is very 'long' and each power set (of an infinite set) is very 'thick.' Hence, any axioms to such effects are in accordance with our intuitive concept. The difficulty is that in order to make specific assertions to increase the length or the thickness, we generally have to use propositions which imply other consequences as well. And then we can no longer justify all these propositions by appealing merely to our (maximum) iterative concept. In particular, there is no known positive principle that guides the search for new axioms to enrich power sets.

For example, consider the hypothesis that the cardinality of the continuum is real-valued measurable. This does deal with the power set of ω, and it decides the continuum hypothesis. But nobody is willing to take it as an axiom. It asserts a special relationship between cardinal numbers defined initially in different ways and is not the kind of proposition which one would be inclined to regard as directly justifiable on the basis of the intuitive iterative concept. Another example is the much more intensively studied axiom of determinacy (*AD*). This axiom *AD* does imply that there are a lot of real numbers (and comparatively few sets of real numbers). But, as it stands, *AD* embodies a generalization of DeMorgan's law with quantifiers (e.g. $\neg\forall x\exists y$ is equivalent to $\exists x\forall y\neg$) to the infinite case and asserts the existence of winning strategies for infinite games.

In its full generality, *AD* contradicts the axiom of choice. The attention is, therefore, mostly concentrated on restricted forms of *AD*. And according to the criterion of pragmatic success, the projective *AD* (say) performs very well indeed. It yields uniform and elegant proofs that all projective sets are Lebesgue measurable and have the Baire property, it yields pleasing new results on the reduction principles, it is easy to state and understand, etc. However, *AD* is not taken as an axiom in the sense that we can see directly from our intuitive concept that it or certain restricted forms of it are true. Rather it is generally viewed as an efficient hypothesis which yields elegant consequences and, in various restricted forms, may be derivable from more intuitive principles ('axioms of infinity' or large cardinal axioms) about the length of the collection of all ordinal numbers.

If we have somehow got hold of the 'real' power set of integers, *CH* should already enjoy a definite truth value even though we might not know what the value is. It is an empirical fact that we do not know how to enrich the power set directly by intuitively evident principles. Hence, the current search for new axioms (in particular, for the purpose of settling *CH*) centers around large cardinal axioms. Parallel to the imperfect information regarding the thickness of the power set of ω, our knowledge of the countable ordinals is also very incomplete. For example, the minimal α such that M_α is a model of *ZF* is a countable ordinal about which we have little to say without reference to the system *ZF*. Even if we look at *ZF* directly in an attempt to build a countable model, we are not clear whether and to what degree the circularity of assuming *ZF* to have a model can be avoided.

The fascination with axioms of infinity leads to the reaction: Why just this one jewel? This is undoubtedly connected with the impression that we can find axioms of infinity which mostly appear to be justifiable by an appeal to the inexact iterative concept of set. To begin with, it is commonly believed that the positive notion of continued iterations is sufficient to justify inaccessible numbers and Mahlo numbers. For example, the existence of (strong) inaccessible numbers means roughly just that the totality of sets obtainable by the procedures of set formation embodied in the axioms of *ZF* forms again a set. Hence, these same procedures are applicable to it to yield other new sets (Zermelo 1930: 29–47). Since the iterative concept permits unlimited extensions, the new axioms are seen to be introduced without arbitrariness. Moreover, each of these axioms, under the assumption of its consistency, can be shown to yield new number-theoretic theorems. Hence, they can be defended both on the ground of intrinsic necessity and, to some extent, on the ground of pragmatic success.

Another method of justifying axioms of infinity is by way of the reflection principles. The iterative concept implies that the universe of all sets is very large. When we have expressed certain properties of the universe, we can already find sets which have these properties. In other words, the reflection principle generalizes the relation of inaccessible

numbers to the axioms of *ZF*. It says that, any time we try to capture the universe from what we positively possess (or can express), we fail the task and the characterization is satisfied by certain (large) sets. Such principles have been applied to justify (to derive) the existence of the inaccessible and the Mahlo numbers, as well as almost all axioms of *ZF*.[14] They have also been applied to justify larger cardinals. But, for example, reflection principles of diverse forms which are strong enough to justify measurable cardinals (by way of 1-extendible numbers) no longer appear to be clearly implied by the iterative concept of set.

There used to be a confused belief that axioms of infinity cannot refute the constructibility hypothesis (and therefore even less the continuum hypothesis) since *L* contains by definition all ordinals. For example, if there are measurable cardinals, they must be in *L*. However, in *L* they do not satisfy the condition of being measurable. This is no defect of these cardinals, unless one were of the opinion that *L* is the true universe. As is well known, all kinds of strange phenomena appear in nonstandard models. However, there does remain a feeling that the property of being a measurable cardinal says more than just largeness, although it implies largeness. It is often felt that the existence of measurable cardinals is more problematic than the existence of inaccessible numbers, even if we disregard the fact that the former is much stronger than the latter. In fact, there are different ways of introducing large cardinals. For example, sometimes we introduce large cardinals by singling out properties of ω in relation to the smaller ordinals and say that there exist cardinals greater than ω which have such properties.

Yet large cardinal hypotheses do occupy a preferred place among the candidates for new axioms about sets because in the majority of cases we expect to be able to show that they just make explicit that the iterative model contains ranks R_α for certain large α. Many of these hypotheses are linearly ordered in the sense that, for two hypotheses H_1 and H_2 we can either (1) derive H_1 from H_2 in *ZF* and find (by assuming H_2) a rank R_α which satisfies *ZF* plus H_1 but does not satisfy H_2, or (2) obtain the same results with H_1 and H_2 interchanged.

There are a number of different aspects of mathematics. In two senses, set theory is not sufficiently abstract to serve as foundations of mathematics. It might be said that we have real numbers as a basic datum, and it is less central how reasoning about real numbers is formalized. In another direction, mathematics is interested in abstract structures such as groups and fields which, though involving concepts like that of set, are independent of the detailed structures of our set theory.

[14] Compare Lévy 1960: 223–38; and Bernays 1961: 1–49.

The modern mathematical theory of categories suggests two rather distinct problems. One is whether the self-applicability of categories is essential so that mathematically interesting proofs would not go through under an interpretation of categories as sets or classes (perhaps of different levels or types). The other is whether such interpretation, even if successful in 'substance,' would not be too artificial as a codification of a type of natural mathematical practice.

5. Comparisons with geometry and physics

Set theory has been compared to geometry and to physics. There are different aspects with regard to which the comparison is made: objects of these disciplines (ontology), sources of our knowledge (epistemology), propositions (axioms or hypotheses) and their truth or acceptability (methodology).

The comparison between Euclid's fifth postulate and the continuum hypothesis is far fetched. Nobody proposed to call *CH* an axiom. There is a feeling that not only is the parallel postulate not evident, but the other postulates are also assumptions (together making up an implicit definition) of which we do not have sharp enough intuition to give justifications. The independence of *CH*, on the other hand, is not accompanied by doubts about the acceptability of the axioms of *ZF*. There is a sense of completeness of geometry with the parallel postulate or its alternatives either as first order theory or as second order theory (with completion coming from the different domain of sets). Hence, even if one takes the position that other postulates are evident or necessary, there is less reason to look for new axioms which would decide the parallel postulate.

Both the parallel postulate and its negation are extensions of 'absolute geometry' (determined by the remaining axioms of geometry) in the weak sense (in the sense of translatability or relative interpretability) (see Gödel 1964: 270–1) [483 in this volume]. This is equally true of *CH* at least relative to *ZF*. But axioms of infinity yield extensions in a stronger sense and there is an asymmetry between an axiom of infinity and its negation. Roughly speaking, an axiom of infinity is stronger and more fruitful than its negation. Epistemologically, there is of course also the difference that geometry is more directly connected to the physical world than set theory.

It seems clear that admitting space as a pure form of intuition need not commit us to the a priori character of Euclidean geometry. For example, we are willing to admit as a consequence of our form of intuition that all physical objects have spatial extension, but then it may be argued that

such a statement is analytic. The scepticism over the parallel postulate is often attributed to the difficulty of envisaging the infinite extension of the straight line. As is well known, there are various equivalent statements which do not mention the infinite extension and can be tested by experiments subject to inevitable inaccuracies in measuring continuous quantities. Two answers can be given to attempts to determine the truth of Euclidean geometry by empirical observations. One would be Poincaré's 'conventionalism.' The other would be to speak of a (local) 'space of intuition' which necessarily satisfies Euclidean geometry. It is true that this second reply is not refuted by experience. Only it is hard to give a convincing positive argument that we do have such exact intuitions, with or without the parallel postulate. Perhaps statements such as 'there are three noncollinear points' are evident and necessary.

It seems curious that while certain obvious things are proved in an elaborate manner, many other gaps are left wide open in Euclid. One explanation might be that the metaphorical definitions of points, lines, etc. are implicitly appealed to. On the other hand, there is a body of central theorems and proofs which are presumably proved quite exactly, and the foundations were accepted with a good deal of tolerance.

Another apparently puzzling feature is the relation of the fifth postulate to the rest. If one doubts the fifth postulate and it is shown to be independent, then the natural conclusion would seem to be that the fifth postulate is not a priori but the others are not affected. Instead, one began to question the necessity of all the postulates. This historical fact is perhaps a combination of two different factors. One is led to realize there is a difficulty in understanding the primitive concepts (or their definitions as originally given). Also, the existence of consistent alternatives shows that we do not have a 'complete' system that is necessary, and that, therefore, we probably do not have enough intuition to justify even parts of the system.

What is gained by comparing set theory with physics? One reason may be the suggestion of mathematical objects. In this comparison, set theory is quite different from arithmetic where, unlike evidence in physics, the general rules such as mathematical induction are perfectly evident.

It has been argued that just as physical objects are natural and necessary for organizing our physical experience, mathematical objects are natural and necessary for organizing our mathematical experience. Physical objects and not merely sensations are immediately given since they are not mere combinations of sensations and our thinking cannot create qualitatively new elements. Or perhaps what is given is not the physical objects but merely something different from sensations which generates the unity of one object out of the diversity of its many aspects. But then one is inclined to think of this something as contributed by the mind. The operation 'set of' is undoubtedly an instrument of synthesis. But there is always lurking somewhere the problem of infinity. To the extent the operation 'set of' suggests a synthesis, we seem to call up a picture of images which can possibly operate on infinite totalities and even permit infinite iterations. But the image of all subsets of a given infinite set (say ω) seems to involve an especially big jump. If somehow we have these subsets, we can apply the operation 'set of' to them. Yet to arrive at all subsets (say of ω), we seem to use something like an analogy.

Perhaps the fact that we can operate with such collections (e.g. in Cantor's diagonal argument) and even use them to prove theorems about natural numbers may be regarded as evidence. But such data are ambiguous and open to interpretations compatible with, e.g., predicative set theory. It seems reasonable to assert that not all data need be associated with certain things acting on our sense organs. But the only other possible source would seem to be either mind or a subtler form of objective reality which either is different from the physical world (e.g. remembering a platonic world) or is the same world but only acts on us in a different way (perhaps by an 'abstract perception').

It seems relatively easy to accept that axioms of set theory force themselves upon us as being true, or even that we feel we have an intuition that produces not only these axioms but an open series of extensions of them. It is perhaps reasonable to assume that all possible extensions will converge. But it seems a stronger assumption to say that the extensions will eventually yield in some sense a unique model so that the continuum hypothesis will be true or false in that model.

To say that mathematical objects exist objectively or even more that in some indeterminate sense we 'perceive' them seems to be stronger yet. Of course, if this is true, we are entitled to use the predicate calculus and reach the conclusion that every statement of set theory is meaningful. For our knowledge, we may still have the problem of recognizing whether a statement is true. This may be the reason why one could believe this strong position and yet not regard the criterion of pragmatic success as entirely superfluous.

The comparison with physics suggests that we look for evidence indirectly through consequences. While it is not necessary that we should be omniscient with regard to our intellectual creations, the difficulty in recognizing new axioms does appear more compatible with the view that mathematical objects exist independently of us. It would seem inevitable that applying this criterion would cost mathematical axioms much of their 'absolute certainty.' For it cannot be denied that success is a matter of degree and consequences (especially lower level consequences which

are the more important for this criterion) do not at all determine the axioms in any unique manner. This does not necessarily obliterate the difference between physics and set theory (as with geometry), at least insofar as we do not at present envisage testing axioms of set theory by consequences in physics.

We may wish to compare the continuum hypothesis with a physical hypothesis which cannot be decided yet. But the analogy is certainly unclear since the latter is not only related to laws (axioms) of physics but depends on a good deal of empirical data. In fact, whatever undecidable propositions in physics may mean, they would seem to be of a radically different nature from those in formal mathematical systems.[15]

6. Digression on unbounded quantifications

The central problems of the iterative concept are: (1) the power set operation (i.e. what subsets are allowed); (2) ordinal numbers (i.e. what ordinals are allowed). Both involve an element of unlimited generality which cannot be rendered completely explicit. An explication of the concept of definite property in the principles of subset formation and replacement is relevant to approximations to this element of generality. In addition, there are principles for generating ordinals which depend on, besides iteration, also analogy and reflection.

Several people have questioned the legitimacy of using unbounded quantifiers (ranging over all sets) in defining new sets, even accepting the iterative interpretation of set theory. It has been suggested, among other things, that we are only justified in using all axioms of *ZF* if we combine it with the intuitionistic predicate calculus. This leads to an exact formal system and mathematical problems concerning the strength of such a system. The philosophical point at issue is, however, by no means equally clear.

If the totality of all sets is an unfinished totality, there is a problem in the use of unbounded quantifiers in the axiom of replacement and in the logical inferences. At any rate the logical notion of sentence is quite alien to Cantor's conception and there is a problem of comparing the sentential generating principle with the conceptual one (less correctly, the arithmetic one).

[15]Gödel points out that the hypothesis of measurable cardinals may imply more interesting (positive in some yet to be analyzed sense) universal number-theoretical statements beyond propositions such as the ordinary consistency statements such as, for instance, the equality of p_n (the n-th prime number) with some easily computable function. Such consequences can be rendered probable by verifying large numbers of numerical instances. Hence, the difference with the hypothesis of the expanding universe is not as great as we may think at first.

It does not seem unreasonable to regard at the same time ω as an unfinished totality and yet allow both quantifiers over all members of ω and the unrestricted law of excluded middle. Here it is clear that every member of ω is reachable from 0 by the successor function. Similarly, if we consider a theory of the second number class, we would be willing to do the same, even though we cannot explicitly give the operations for generating new countable ordinals. From this point of view, the current practice with the universe V of all sets seems entirely consistent.

It might be argued that we do look beyond ω_1 while V is a sort of absolute limit; ω_1 is a set but V is not a set. But the situation does not change, if we are only interested in countable ordinals and regard ω_1 as our universe. And the crucial point is, we do not anticipate incompatible alternative extensions of various approximations to ω_1 or V. Even though there are different ways of approximating to ω_1 and to V, we expect them to converge.

The genetic element is tied up with the distinction between viewing a class as one and as many. It is not excluded that we can talk about all sets and even form classes by conditions on all sets. Only in order that class be treated as one, we have to build it up from below. This imposes a requirement on its members (they must be 'given') but not on how one is to select sets from all given sets to form a new set. This is typically the situation with the replacement axiom: we make sure that all members of the new set are sets but choose them by arbitrary means, including the use of unbounded quantifiers. In particular, only a class as one can be a member of other classes or sets.

The concept of an unfinished or subjectively unfinishable totality is distinguishable from the idea that existence must agree with provable generation according to predetermined generating operations. Rather it permits the classical interpretation of quantifiers. We may also appeal to the reflection principles to argue that the unbounded quantifiers are not really unbounded. We are inclined to say that it is not problematic that we accept the classic logic with regard to propositions about all sets. There is no objection against viewing such propositions as determining definite properties.

If we adopt a constructive approach, then we do have a problem in allowing unlimited quantifiers to define other sets. Even then there remains the possibility of accepting the law of excluded middle. The difficulty is rather in establishing universal conclusions because we cannot survey all permissible operations.

Frege thinks of individuals, predicates, predicates of predicates, etc. He identifies sets with extensions of predicates and treats them as individuals (on the same level with individuals). This suggests immediately

the idea of a type hierarchy of extensions, since extensions of predicates seem to be more closely related to predicates than to individuals. Alternatively, we may wish to think of all sets as objects (individuals) but distinguish them from extensions of predicates. In that case, we would be led to something like a finite type theory (perhaps ramified) based on the current set theory (a theory of individuals). But such a conception has little that is positive to say about how sets are to be generated.

7. Extracting axioms of set theory from Cantor's writings

As before, we discuss Cantor's views by reference to his collected works. It is relatively easy to be dissatisfied with Cantor's philosophical speculations on the transfinite. We can comfort ourselves that there is no need to take Cantor's flight into theology seriously, especially since we now possess more reasonable defenses of set theory.

> One proof proceeds from the concept of God and concludes first from the highest perfection of God's essence the possibility of creating a transfinitum ordinatum. It then goes on to conclude from His divinity and glory the necessity of the actual successful creation of the transfinitum. Another proof shows a posteriori that the assumption of a transfinitum in natura naturata yields a better (because more complete) explanation than the opposite hypothesis, of the phenomena especially of the organisms and the psychical facts (1886, p. 400).

> In one sense we may regard the integers as real insofar as they take up, by dint of definitions, a wholly determined place in our understanding, are well distinguished from all other ingredients of our thinking, but stand in definite relations to them and thereby modify the substance of our spirit in a definite way; I propose to call this kind of reality of the numbers their *intrasubjective* or *immanent reality*. The numbers can, however, also be ascribed reality insofar as they must be considered as an expression or picture of events and relations in the world outside our intellect, as further the different number classes I, II, III, etc. are representatives of cardinalities which actually appear in the physical and spiritual nature. I call this second kind of reality the *transsubjective* or *transient reality* of the integers...
>
> This coherence of both kinds of reality has its proper root in the *unity* of *all, in which we ourselves participate*. The allusion to this coherence serves here the purpose of deriving a consequence for mathematics which seems very important to me, namely that in developing ideas in it we *only* have to account for the *immanent* reality of its concepts, and are not at all obliged to test their *transient* reality (1883, pp. 181–2).

This last special property of mathematics is for Cantor the ground for calling mathematics 'free.'

In the case of Cantor, what is more interesting for philosophy is perhaps not so much his metaphysical speculations as the conceptual framework revealed by his mathematical practice.

Cantor uses multiplicity or manifold (Vielheit) as a primitive concept which corresponds to class in current usage. According to him there are two kinds of (definite) multiplicity: the absolutely infinite or inconsistent multiplicities (proper classes in current usage) and the consistent ones which are called sets. He then states explicitly three axioms (pp. 443–4).

C1 Two equivalent classes are either both sets or both proper classes. (One-one replacement).[16]

C2 Every subclass of a set is a set. (Subset formation).

C3 For any set of sets, the elements of these sets again form a set; in other words, the union of a set of sets is a set. (Union).

There are undoubtedly axioms which appear too obvious (and presumably too numerous) to Cantor to be stated explicitly. Such axioms can of course be added without violating Cantor's intention.

Like most mathematicians, Cantor uses implicitly the axiom of extensionality, for example, in establishing $P=Q$ for two point sets P and Q (but compare the introduction of $\equiv$ on p. 145). By the way, Dedekind mentions this axiom explicitly (1888, §2).

C4 If two classes A and B have the same elements, then $A=B$ (in particular, two sets with the same extension are equal). (Extensionality).

A more interesting case is the power set axiom. In at least two connections, Cantor uses implicitly something like the power set axiom. In defining exponentiation of cardinals (pp. 287–9, 1895) Cantor considers the totality of functions from a set a into a set b and states explicitly that the totality is again a set. In particular, he shows that the cardinality of the linear continuum is $2^{\aleph_0}$. In presenting his diagonal argument to establish $2^\lambda > \lambda$, for any cardinal λ, he considers for every set a (in particular, the set of all real numbers) with cardinality λ, the set b of all functions $f(x)$ which takes only 0 and 1 as values with x ranging over all members of a (in particular, all reals $\geqslant 0$, $\leqslant 1$) (pp. 279–80, 1890–1). Here, b is the power set of a.

C5 For every set a, all its subsets form a set. (Power set).

Of course, Cantor never doubts there are infinite sets. He freely uses the sets of all integers, all algebraic numbers, all reals, etc. (pp. 115, 126,

[16] Cantor does seem to apply implicitly the axiom of replacement to get ω_ω from $\omega, \omega_1, \ldots$ in his 1895 paper (p. 296).

143, etc.). More specifically, he states explicitly (p. 293, 1895) that the totality of all finite cardinals forms a set.

C6 There is a set containing 0, 1, 2, etc. (Infinity).

A basic indefiniteness in the above axioms is the notion of equivalence or one-one correspondence in C1: what language forms are permissible in specifying the one-one correspondence? The problem is illustrated by Zermelo's discussion (1908: §1.4, §1.6) of definite properties (which are presumably contrasted with things like poetic images and theological characterizations). One expects Cantor to accept a broad range of one-one correspondences so that, for example, any formula in the second order language with set and class variables would be permissible; or alternatively one might choose to exclude bound class variables.

As far as I know, Cantor does not discuss well-founded sets. But I believe he operates under the assumption that all sets are well-founded.

It is well known that Cantor does not consider the axiom of choice and often uses it implicitly. For example, on p. 293 (1895), he gives a 'proof' that every infinite set contains a subset of cardinality $\aleph_0$ but is not aware that an appeal to the axiom of choice is needed.

Cantor is much interested in the well-ordering theorem in the form that (1) the totality of all cardinal numbers can be well-ordered (p. 280, 1890–1; p. 295, 1895) or that (2) all infinite cardinal numbers are alephs (p. 447, 1899). It is perhaps worth remarking that on p. 280 and p. 295 he speaks of the totality of all cardinals as a 'well-ordered set' in quotation marks. Undoubtedly the later distinction between sets and classes is a way of removing these quotation marks. Actually, on p. 280 Cantor claims to have proved (1) in his paper of 1883 (pp. 165–208). It seems that this erroneous assertion is based on the implicit assumption of (2) as evident or proven. On p. 285 (1895), Cantor also asserts the comparability of any two cardinal numbers with a promise to return to a proof of this.

On p. 447, Cantor attempts to prove (2) by arguing that the cardinality of every set is an aleph. Suppose V is a class with no aleph as its cardinal number. Then, Cantor argues, the well-ordered class On of all ordinals is projectable into the class V and there must exist a subclass of V which is equivalent to On. Hence, by C1 and C2, V must be a proper class. The claim that labelling distinct members of V would use up all members of On appeals to some form of the axiom of choice.

From the above discussion, it would appear that what is currently called ZF or perhaps the second order theory of ZF would be a reasonable codification of Cantor's concept of set. This is under the assumption that we identify ordinal and cardinal numbers with sets in the now familiar manner and that one-one correspondences be specified more explicitly. Also, Zermelo's formulation of the axiom of choice and his use of it in proving the well-ordering theorem is a definite advance beyond Cantor.

It would seem a relatively simple matter to introduce the rank hierarchy and observe that every set has a rank. Cantor, however, seems to consider primarily only hierarchies of all numbers but not those of all sets. Mirimanoff seems to be the first to formulate explicitly the iterative concept with the rank hierarchy (1917a, b).

8. The hierarchies of Cantor and Mirimanoff

Cantor considers the well-ordered classes of all ordinals and all infinite cardinals (p. 444):

A $0, 1, 2, \ldots, \omega, \ldots$
B $\aleph_0, \aleph_1, \aleph_2, \ldots, \aleph_\omega, \ldots$

In the current treatment, each member of *A* or *B* is a set and, in general, $\aleph_\alpha$ is identified with ω_α ($\aleph_0 = \omega$). The consideration of the exponentiation of infinite cardinals (p. 288) and the cardinality of the power set of an infinite set (p. 280) also suggests another well-ordered class:

C $\aleph_0, 2^{\aleph_0}, 2^{2^{\aleph_0}}, \ldots$, ($C_0 = \aleph_0$, $C_{\alpha+1} = 2^{C_\alpha}$, $C_\lambda =$ the union of all C_β, $\beta > \lambda$, λ a limit number).

The generalized continuum hypothesis says in effect that each term of *B* is identical with the corresponding term of *C*. The continuum hypothesis asserts that $2^{\aleph_0} = \aleph_1$. It is well known that Cantor introduced the continuum hypothesis and spent much effort trying to prove it. As early as 1878, Cantor asked the question as to how many classes would result if we divide all infinite point sets into different classes according to their cardinalities, and stated (p. 132): 'By an inductive procedure, the presentation of which we do not enter upon here, one can prove the theorem that the number of classes of linear point sets yielded by this principle of partition is a finite one and, in fact, it is *two*.' And then in 1883, Cantor asked for the cardinality of the continuum and stated (p. 192): 'I hope to be able to answer soon with a strict proof that the cardinality sought is none other than that of our *second number class*.' This is followed by a footnote (number 10, see p. 207) stating in effect that $2^{\aleph_1} = \aleph_2$. At the end of his most important paper (published 1879–84), he promised again (p. 244, 1884) that the continuum hypothesis 'will be proved in later sections.'

Cantor's continuum problem is a sharp formulation of the simple and

intrinsically interesting question: how many points are there on a line or how many sets of integers are there? It is remarkable that Cantor is able not only to give a stable extension of the concept of (cardinal) number to infinite sets without arbitrariness but also to give the well-ordered class of all alephs as a basis for comparison. The class of alephs uses the class of ordinal numbers and Cantor's concept of number classes to get the next aleph after each given single aleph or sequence of alephs indexed by an ordinal. That all infinite cardinals are alephs (a form of the well-ordering theorem) depends on the axiom of choice. Hence, Cantor's great interest in the well-ordering theorem is easily understandable, seeing that it is needed to give a definite shape to the collection of all cardinal numbers. The very fundamental character of the continuum problem as well as the impressive achievements of Cantor in arriving at a sharp formulation of the problem also explain both Cantor's organization of his ideas and his obsession with the continuum hypothesis. A solution of the continuum problem would have been the crowning event of his whole intellectual development.

Cantor had obtained the Burali-Forti paradox at least two years before Burali-Forti's publication in 1897 and communicated it to Hilbert in 1896 (p. 470). He evidently did not find the phenomenon shocking and his distinction between sets and inconsistent multiplicities does not strike one as an ad hoc device but rather like a sharpening of an incomplete intuition quite in the spirit of his general approach. But his description of the distinction is admittedly vague and all too brief. 'For a multiplicity can be such that the assumption of a "being together" of *all* of its elements leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as "one finished thing". Such multiplicities I call *absolutely infinite* or *inconsistent multiplicities*' (p. 443). These letters to Dedekind were not published until 1932.

In 1917, Mirimanoff published two papers[17] in which ideas similar to Cantor's are discussed in great detail and pursued further.

With regard to ideas familiar today, the concept of well-founded sets is introduced along with the rank function and employed to show (all in M1) that every well-founded set has a rank (the iterative model). Furthermore, the representation of ordinals commonly associated with the name of von Neumann is proposed in M1 and developed more extensively in M2. It is perhaps of some historical interest that von Neumann, in his paper of 1925 (219–40), did refer to M1 in connection with well-founded sets (1961–3, 1: 46n.) but apparently was not aware of the anticipation of his definition of ordinals.

[17]Mirimanoff 1917a: 37–52; 1917b: 207–17; 1920: 29–52. These will be referred to respectively as M1, M2, M3.

The fundamental problem of M1 is (p. 38): What are the necessary and sufficient conditions for a set of individuals to exist? The distinction between existent and nonexistent sets corresponds to that between sets and proper classes. For linguistic convenience, we shall translate Mirimanoff's terms into the familiar ones. As another example of different terminology, he speaks of 'ordinary sets' instead of 'well-founded classes (or sets).'

In M1, each class is associated with a membership tree that goes from the class to all of its members, and thence to all their members, etc. (p. 41). Each path in this tree is called a descent and a class is well founded if all the descents in its tree are finite (p. 42). The trees can only stop at indecomposable elements which are the noyaux (p. 43, nuclei, the urelements). In particular, the empty set is taken as an urelement and denoted by e. The well-founded sets are suggested by the paradox of sets which are not their own members. It is clear that the class of all well-founded sets is not a set (p. 43). This approach brings out a common feature of Russell's and Burali-Forti's paradoxes.

It is not assumed that we are only interested in well-founded sets, but rather a solution for the fundamental problem is only proposed for well-founded sets. Six axioms are stated in M1.

- P1 Subset formation. Every subclass of a set is a set (p. 44. This is 'a property of sets which is far from being evident but which I regard as true, at least for the sets I consider in this work,' p. 43).
- P2 A class equivalent to the class *On* of all ordinals is not a set (p. 45. By P1, any class containing a subclass equivalent to *On* is not a set).
- P3 The urelements form a set which is regarded as given or known (p. 48).
- P4 Power set. For every set a of well-founded sets, there is a set of all subsets of a (p. 49).
- P5 Union. For every set a of well-founded sets, there is a set of all members of members of a (p. 49).
- P6 One-one replacement. Given a set a and a one-one correspondence correlating each member of a with a well-founded set, there is a set of all the images of members of a (p. 49).

The concept of rank is introduced explicitly (p. 51): the rank of a well-founded set is the smallest ordinal number greater than the ranks of its members. The rank of an urelement (in particular, of e) is zero.

Two theorems are then proved (p. 51).

> Th. 1 Every well-founded set has a rank. It is first observed that a set has a rank if all its members have ranks. This is proved by a lemma to the effect that every set of ordinal numbers has a rank. Since ordinals under the Cantor conception are not sets, the von Neumann ordinals are introduced to represent them and P6 is applied to obtain the lemma (p. 50). If now a well-founded set x did not have a determinate rank, then there is at least one member x_1 of x having the same property; similarly x_1 has at least a member x_2 not having a determinate rank, and so forth – an absurd result, seeing that the whole sequence $x, x_1, x_2, \ldots$ stops at an urelement whose rank is zero.

This elegant proof makes an implicit use of the countable axiom of choice and is familiar nowadays.

> Th. 2 For every α, the collection R_α of all well-founded sets of rank α forms a set.

First, if the theorem is true for all $\alpha < \pi$, then it is also true for π. Thus let Σ be the union of R_α for $\alpha < \pi$. By P6 and P5, this union is a set. But R_π is a collection of subsets of Σ and is therefore a set by P4 and P1. But the class of all ordinals is well-ordered and the class of ordinals for which Th. 2 is false, if not empty, must have a least number.

The two theorems together yield the result: there is a relation $S(\alpha, y) \equiv y = R_\alpha$, and for all well-founded sets x, $\exists\alpha\exists y(S(\alpha,y) \wedge x \in y)$. Let $H = \hat{x}(\exists\alpha)(S(\alpha,y) \wedge x \in y)$.

The solution for the fundamental problem thus obtained in M1 is the following (pp. 51–2): a collection of well-founded sets is a set if and only if the ranks of its members are bounded above by an ordinal number.

Clearly if we assume the axiom of foundation (that all sets are well founded), we can delete from the above discussion the qualification of well-foundedness, and obtain the result that the universe is the same as the union of all R_α (briefly $V = H$) and therewith reach the iterative concept of set described in §1.

Intuitively it seems very plausible that since we collect 'given' objects to form new sets, there should be enough ordinals to index the stages of this process of iterated expansion. The axiom of foundation sharpens the concept of iteration. Given the axiom, which says that every path in a membership tree leads back to an urelement in a finite number of steps, it becomes more plausible that we can, beginning with the set of all urelements, reach each set by steps indexed by ordinal numbers. The possibility of deriving $V = H$ from the axiom of foundation of course shows that no strong new axiom can contradict the iterative concept without refuting some of the basic axioms commonly accepted. In this respect, $V = H$ is very different from the much stronger 'axiom' of constructibility.

Of course, the basic axioms for getting larger sets are replacement and power set. By definition, power set only increases the rank by 1, and replacement has also this 'local' character because every set of ordinal numbers has an upper bound which is again an ordinal number. Replacement itself assures that the indices of the members of the image set again form a set. A justification of the axiom would be that proper classes are so large that a one-one correspondence never gets from a set to a proper class.

Once we have $V=H$, it seems reasonable to strengthen Cantor's axiom C1 to say also that (C1*) all proper classes are equivalent to V. Thus, given any class, either there is a bound on the ranks of all its members and then it is a set, or else the ranks are unbounded and then it forms a proper class C. If we use the members of C to count the members of V rank by rank, then we cannot stop at any R_α because we would then have a one-one correspondence between a proper class C and a set which is the union of $R_0, \ldots, R_\alpha$. To carry out this argument, we have to assume the global axiom of choice that the universe V can be well-ordered because the well-ordering of each R_α is not given explicitly by the local axiom of choice. Given $V=H$, this appears to be the natural generalization of the usual local axiom of choice by which each R_α can be well-ordered. The axiom C1* was first introduced by von Neumann (1925). The axiom of foundation and the rank model were rediscovered and treated more thoroughly by von Neumann (1925 and 1929) and also by Zermelo (1930). A particularly nice formulation of the axiom of foundation would seem to be: if every subset of a class X belongs to X, then $X=V$.

In M1, the ordinal numbers, 1, 2, 3, etc. are represented by $\{e\}$, $\{e,\{e\}\}$, $\{e,\{e\},\{e,\{e\}\}\}$, etc. These are obtained as follows. Let x be a well-ordered set, and let y be the set of all its segments, including the segment e. Replace these segments in y by the sets of the segments of these segments and apply an analogous transformation to the segments introduced in this manner, and so on (p. 45). A definition not appealing to given well-ordered sets is also given (p. 47): Definition of *On*: a set x represents an ordinal if: (1) x is a well-founded set based on the urelement e; (2) if y and z are two distinct elements of x, then $x \in y$ or $y \in x$ (connected); (3) $y \in x$, then $y \subseteq x$ (transitive).

If we adopt the axiom of foundation, this of course yields the currently employed definition of *On* which is usually regarded as an improvement over von Neumann's rediscovery of 1923 (199–208). Further properties of *On* are derived in M2. In the conclusion of M2, it is pointed out that *On* can be used in place of Cantor's ordinal numbers in dealing with well-ordered sets. 'I do not know whether this indirect method presents real advantages. In any case the classic theory of Cantor appears thus under a new aspect' (p. 217).

A totally new problem in the foundations of set theory is considered in M3, namely the nature of definite properties implicitly employed in stating the axioms of subset formation and replacement. The discussions are not nearly as conclusive as the extensional considerations in M1.[18]

Turning to the origins of Cantor's development of set theory we append some historical notes on a few anticipations and independent discoveries of some of Cantor's ideas in set theory. As is well known, Cantor began his mathematical career by works on the trigonometric series.[19] In trying to extend the uniqueness of representation in terms of these series to certain functions with infinitely many points of discontinuity, Cantor was faced with the problem of singling out suitable infinite sets of points on the line. This led to the notion of the derived set P' of a set P (viz. the set of limit points of P) which not only marked the beginning of Cantor's study of point set theory but paved the way for his construction of transfinite ordinal numbers later on.

In 1872, Cantor considered finite iterations of the operation of derived set and observed that there can be point sets P such that, for every n, $P^{(n)}$ is not empty and hence is infinite (p. 92). In 1880 (p. 145), Cantor introduced infinite iterations to ∞, $\infty+1$, $n_0\infty^m + \ldots + n_m$, ∞^∞, etc. In particular, $P^{(\infty)}$ is, for example, identified with the intersection of $P', P^{(2)}$, etc. 'Here we see a dialectical generation of concepts, which always leads yet farther, and remains both free from every arbitrariness and necessary and logical in itself' (p. 148). This was followed in 1883 by the first extensive development of transfinite numbers. Cantor observed that further progress of his investigations would depend on an extension of the concept of number which nobody had attempted yet, and that without this extension of the concept of number:

> It would be impossible for me to advance freely a single step in the theory of sets; in this circumstance a justification or, if necessary, an excuse, may be found for my introduction of apparently strange ideas in my considerations. These concern an extension or continuation of the sequence of integers into the infinite; however daring these may appear, I can nevertheless express not only the hope but the firm conviction that this extension will in time be regarded as thoroughly simple, proper, and natural. At the same time, I by no means conceal from myself the fact that with this enterprise I place myself in a certain opposition to widespread views on the mathematical infinite and to oft-defended opinions on the essence of number (p. 165).

In a paper of 1880 (pp. 115–28), P. du Bois-Reymond claimed priority

[18] For related historical matters, compare also part II of the survey of Skolem's work in logic (1970: 35–40).

[19] A long discussion of the relation of Cantor's work in this area to his predecessors' is contained in a series of papers by Jourdain (1906–13).

on the concept of derived sets of infinite order. He considered a set D of intervals so distributed on the straight line that every interval contains some member of D as a part. Then he asserted: 'We are led to this kind of distribution of intervals, of which I have several examples, when we seek for points of accumulation of order ∞, whose existence I indicated to Mr. Cantor of Halle by letter years ago.' He also mentioned that he had introduced his notion of 'pentachic' before Cantor did his equivalent notion of 'everywhere dense.'

A more interesting anticipation was du Bois-Reymond's use of the diagonal method in his theory of growth[20] nearly twenty years before Cantor published in 1892 his famous diagonal proof of the theorem that every set has more subsets than elements. Consider increasing functions of one real variable x, for $x>0$. For two such functions $f(x)$ and $g(x)$, du Bois-Reymond stipulated that $f<g$ if $f(x)/g(x)$ tends to 0 as x increases indefinitely. The following theorem was proved. Let $f_1, f_2, \ldots$ be any sequence of increasing functions such that $f_1<f_2<f_3, \ldots$, then there exists an increasing function f such that $f_n<f$, for all n. He defined a new sequence $g_1, g_2, \ldots$ such that $g_1=f_1$, and $g_{n+1}(x)>g_n(x)$, for all x. Thus, by hypothesis, there exists x_1 such that $f_2(x)>g_1(x)$, for $x>x_1$. Let $g_2(x)=f_2(x)$, for $x>x_1$; $g_2(x)=g_1(x)+1$, for $x\leqslant x_1$. Similarly, let $g_{n+1}(x)=f_{n+1}(x)$, for $x>x_n$; $g_{n+1}(x)=g_n(x)+1$, for $x\leqslant x_n$. The desired function f can then be defined: for every n, $f(n)=g_n(n)$ and $f(x)=g_{n+1}(x)$; for $n<x<n+1$. This theorem is analogous to Cantor's theorem that every fundamental sequence of ordinal numbers defines a greater ordinal. An analog of the immediate successor would be to go from f to g such that $g(x)=xf(x)$.

[20] The relevant papers by P. du Bois-Reymond are: 1869: 10–45 (especially §7); 1873: 61–91 (especially the appendix); 1875: 363–414, and 1877: 149–67.

Bibliography

Ackermann, W. 1940. "Zur Widerspruchsfreiheit der reinen Zahlentheorie." *Mathematische Annalen*, vol. 117.

Alston, W. P. 1958. "Ontological Commitments." *Philosophical Studies*, vol. 9. Reprinted in Benacerraf and Putnam 1964.

Ambrose, A. 1935. "Finitism in Mathematics," parts 1 and 2. *Mind*, n.s. vol. 44.

1955. "Wittgenstein on Some Questions in the Foundations of Mathematics." *Journal of Philosophy*, vol. 52.

1956. "On Entailment and Logical Necessity." *Proceedings of the Aristotelian Society* (1955–6), n.s. vol. 56.

Anderson, A. R. 1958. "Mathematics and the 'Language Game.'" *Review of Metaphysics*, vol. 11. Reprinted in Benacerraf and Putnam 1964.

Anscombe, G. E. M. 1959. *An Introduction to Wittgenstein's Tractatus*. London: Hutchinson.

Ayer, A. J. 1936a. "The *A Priori*." In Ayer 1936b. Excerpted in Benacerraf and Putnam 1964 and reprinted in this volume.

1936b. *Language, Truth and Logic*. London: Gollancz; New York: Oxford University Press; 2nd ed., Gollancz, 1946; reprinted, New York: Dover, 1946.

1951. "Symposium: On What There Is II." In *Freedom, Language, and Reality. Aristotelian Society Supplementary Volume 25*.

1959. Ed. *Logical Positivism*. New York: Free Press.

Bachmann, H. 1955. "Transfinite Zahlen." *Ergebnisse der Mathematik und ihrer Grenzgebiete*, n.s. vol. 1.

Barwise, J. 1971. "Infinitary Methods in the Model Theory of Set Theory." In *Logic Colloquium '69. Proceedings of the Summer School and Colloquium in Mathematical Logic* (Manchester, England, 1969). Edited by R. O. Gandy and C. M. E. Yates. *Studies in Logic and the Foundations of Mathematics*, vol. 61. Amsterdam: North-Holland.

1975. *Admissible Sets and Structures*. New York: Springer.

1977. Ed. *Handbook of Mathematical Logic. Studies in Logic and the Foundations of Mathematics*, vol. 90. Amsterdam and New York: North-Holland.

Baumann, J. 1868–9. *Die Lehren von Raum, Zeit und Mathematik in der neueren Philosophie nach ihrem ganzen Einfluss dargestellt und beurtheilt*. 2 vols. Berlin: Reimer.

Becker, O. 1954. *Grundlagen der Mathematik in geschichtlicher Entwicklung*. Freiburg: Alber.

Behmann, H. 1934. "Sind die mathematischen Urteile analytisch oder synthetisch?" *Erkenntnis*, vol. 4.

Benacerraf, P. 1965. "What Numbers Could Not Be." *Philosophical Review*, vol. 74. Reprinted in this volume.

1967. "God, the Devil, and Gödel." *The Monist*, vol. 51.

1973. "Mathematical Truth." *Journal of Philosophy*, vol. 70. Reprinted in this volume.

Benacerraf, P., and Putnam, H., eds. 1964. *Philosophy of Mathematics: Selected Readings*. Englewood Cliffs, N.J.: Prentice-Hall.

Bennett, J. 1959. "Analytic-Synthetic." *Proceedings of the Aristotelian Society* (1958–9), n.s. vol. 59.

1961a. "A Myth about Logical Necessity." *Analysis*, vol. 21.

1961b. "Symposium: On Being Forced to a Conclusion I." *Aristotelian Society Supplementary Volume 35*.

Bernays, P. 1935. "Sur le platonisme dans les mathématiques." *L'Enseignement mathématique*, 1st ser. vol. 34. Translated as "On Platonism in Mathematics" by C. D. Parsons in Benacerraf and Putnam 1964 and reprinted in this volume.

1937–54. "A System of Axiomatic Set Theory," parts 1–7. *Journal of Symbolic Logic*, vols. 2, 6, 7, 8, 13, and 19.

1941. "Sur les questions méthodologiques actuelles de la théorie hilbertienne de la démonstration." *Les Entretiens de Zurich sur les fondements et la méthode des sciences mathématiques* (1938). Edited by F. Gonseth. Zürich: Leemann.

1957. "Von der Syntax der Sprache zur Philosophie der Wissenschaften." *Dialectica*, vol. 11.

1958. *Axiomatic Set Theory*. With a historical introduction by Abraham A. Fraenkel. Amsterdam: North-Holland.

1959. "Comments on Ludwig Wittgenstein's *Remarks on the Foundations of Mathematics*" (Wittgenstein 1956). *Ratio*, vol. 2. Reprinted in Benacerraf and Putnam 1964.

1961. "Zur Frage der Unendlichkeitsschemata in der axiomatischen Mengenlehre." In *Essays on the Foundations of Mathematics, Dedicated to A. A. Fraenkel on His Seventieth Anniversary*. Edited by Y. Bar-Hillel, E. I. J. Poznanski, M. O. Rabin, and A. Robinson. Jerusalem: Magnes Press, Hebrew University.

Beth, E. W. 1956. "Semantic Construction of Intuitionistic Logic." *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Afdeeling Letterkunde*, n.s. vol. 19.

1957. *La Crise de la raison et la logique*. Paris: Gauthier-Villars.

1959. *The Foundations of Mathematics*. Amsterdam: North-Holland.

Birkhoff, G., and MacLane, S. 1941. *A Survey of Modern Algebra*. New York: Macmillan.

Bishop, E. 1967. *Foundations of Constructive Analysis*. New York: McGraw-Hill.

Black, M. 1933. *The Nature of Mathematics: A Critical Survey*. London: K. Paul, Trench, Trubner; New York: Harcourt, Brace; reprinted, London: Routledge and Kegan Paul, 1950.

1948. "The Semantic Definition of Truth." *Analysis*, vol. 8.

Blumenthal, L. 1940. "A Paradox, a Paradox, a Most Ingenious Paradox." *American Mathematical Monthly*, vol. 47.

Bolzano, B. 1851. *Paradoxien des Unendlichen*. Leipzig: Reclam. Translated as Bolzano 1950.

1950. *Paradoxes of the Infinite*. A translation of Bolzano 1851 by F. Přihonský, with a historical introduction by D. A. Steele. London: Routledge and Kegan Paul; New Haven: Yale University Press.

Boole, G. 1847. *The Mathematical Analysis of Logic, Being an Essay Toward a Calculus of Deductive Reasoning*. Cambridge, England: Macmillan, Barclay, and Macmillan; reprinted, Oxford: Blackwell; New York: Philosophical Library, 1948.

1854. *An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton and Maberly; reprinted, New York: Dover, 1951.

Boolos, G. 1971. "The Iterative Conception of Set." *Journal of Philosophy*, vol. 68. Reprinted in this volume.

1975. "On Second-Order Logic." *Journal of Philosophy*, vol. 72.

Borel, E. 1912. "La Philosophie mathématique et l'infini." *Revue du mois*, vol. 14.

Braun, S., and Sierpiński, W. 1932. "Sur quelques propositions équivalentes à l'hypothèse du continu." *Fundamenta Mathematicae*, vol. 19.

Bridgman, P. W. 1934. "A Physicist's Second Reaction to Mengenlehre." *Scripta Mathematica*, vol. 2.

Brouwer, L. E. J. 1907. *Over de grondslagen der wiskunde*. Doctoral thesis, Municipal University of Amsterdam, 1907. Amsterdam and Leipzig: Mass and van Suchtelen; Groningen: Noordhoff.

1908. "Die onbetrouwbaarheid der logische principes." *Tijdschrift voor wijsbegeerte*, vol. 2. Reprinted in Brouwer 1919.

1909. "Die möglichen Mächtigkeiten." In *Atti del IV Congresso Internazionale dei Matematici* (Rome, 1908), 3 vols. Edited by G. Castelnuovo. Rome: Salviucci.

1912. *Intuitionisme en Formalisme*. Groningen: Noordhoff. Reprinted in Brouwer 1919. Translated as Brouwer 1913.

1913. "Intuitionism and Formalism." A translation of Brouwer 1912 by A. Dresden. *Bulletin of the American Mathematical Society*, vol. 20. Reprinted in Benacerraf and Putnam 1964 and in this volume.

1919. *Wiskunde, waarheid, werkelijkheid*. Groningen: Noordhoff.

1925–7. "Zur Begründung der intuitionistischen Mathematik," parts 1–3. *Mathematische Annalen*, vols. 93, 95, 96.

1927. "Über Definitionsbereiche von Funktionen." *Mathematische Annalen*, vol. 97. Translated as "On the Domains of Definition of Functions" by S. Bauer-Mengelberg in van Heijenoort 1967.

1929. "Mathematik, Wissenschaft, und Sprache." *Monatshefte für Mathematik und Physik*, vol. 36.

1949. "Consciousness, Philosophy, and Mathematics." In *Proceedings of the Tenth International Congress of Philosophy* (Amsterdam, 1948), vol. 1, pt. 2. Edited by E. W. Beth, H. J. Pos, and J. H. A. Hollak. Amsterdam: North-Holland. Reprinted in Benacerraf and Putnam 1964 and in this volume.

1952. "Historical Background, Principles, and Methods of Intuitionism." *South African Journal of Science*, vol. 49.

Burali-Forti, C. 1897. "Una questione su i numeri transfiniti." *Rendiconti del Circolo Matematico di Palermo*, vol. 11. Translated as "A Question on Transfinite Numbers" by J. van Heijenoort in van Heijenoort 1967.

Butts, R. E., and Hintikka, J., eds. 1977. *Logic, Foundations of Mathematics, and Computability Theory. Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science* (London, Ontario, 1975), pt. 1. Dordrecht: Reidel.

Cantor, G. 1883a. "Über unendliche, lineare Punktmannichfaltigkeiten," part 5. *Mathematische Annalen*, vol. 21. Reprinted as Cantor 1883b.

1883b. *Grundlagen einer allgemeinen Mannichfaltigkeitslehre. Ein mathematischphilosophischer Versuch in der Lehre des Unendlichen.* Leipzig: Teubner. Reprinted in Cantor 1932.

1895–7. "Beiträge zur Begründung der transfiniten Mengenlehre," parts 1 and 2. *Mathematische Annalen*, vols. 46 and 49. Reprinted in Cantor 1932. Translated as Cantor 1915.

1915. *Contributions to the Founding of the Theory of Transfinite Numbers.* A translation of Cantor 1895–7 by P. E. B. Jourdain. Chicago and London: Open Court; reprinted, New York: Dover, 1952.

1932. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts.* Edited by E. Zermelo. Berlin: Springer; reprinted, Hildesheim: Olms, 1962.

Carnap, R. 1928a. *Der logische Aufbau der Welt.* Berlin: Weltkreis.

1928b. *Scheinprobleme in der Philosophie; das Fremdpsychische und der Realismusstreit.* Berlin: Weltkreis.

1930a. "Die alte und die neue Logik." *Erkenntnis*, vol. 1. Translated as "The Old and the New Logic" by I. Levi in Ayer 1959.

1930b. "Die Mathematik als Zweig der Logik." *Blätter für deutsche Philosophie*, vol. 4.

1931. "Die logizistische Grundlegung der Mathematik." *Erkenntnis*, vol. 2. Translated as "The Logicist Foundations of Mathematics" by E. Putnam and G. J. Massey in Benacerraf and Putnam 1964 and reprinted in this volume.

1934a. "Die Antinomien und die Unvollständigkeit der Mathematik." *Monatshefte für Mathematik und Physik*, vol. 41.

1934b. *Logische Syntax der Sprache.* Vienna: Springer. Translated as Carnap 1937.

1935a. "Formalwissenschaft und Realwissenschaft." *Erkenntnis*, vol. 5. Translated as "Formal and Factual Science" in Feigl and Brodbeck 1953.

1935b. "Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik." *Monatshefte für Mathematik und Physik*, vol. 42.

1935c. *Philosophy and Logical Syntax.* London: K. Paul, Trench, Trubner; reprinted, New York: AMS Press, 1979.

1937. *The Logical Syntax of Language.* A translation of Carnap 1934b by A. Smeaton. London: K. Paul, Trench, Trubner; New York: Harcourt, Brace; reprinted several times.

1939. *Foundations of Logic and Mathematics. International Encyclopedia of Unified Science*, vol. 1. Chicago: University of Chicago Press.

1942. *Introduction to Semantics.* Cambridge: Harvard University Press.

1943. *Formalization of Logic.* Cambridge: Harvard University Press.

1947. *Meaning and Necessity.* Chicago: University of Chicago Press; 2nd ed., enlarged, 1956. Sections 2 and 3 are reprinted in Feigl, Sellars, and Lehrer 1972.

1950. "Empiricism, Semantics, and Ontology." *Revue internationale de philosophie*, vol. 4. Reprinted in Carnap 1947 (2nd ed.), Linsky 1952, and Feigl, Sellars, and Lehrer 1972. The slightly modified version in *Meaning and Necessity* was published in Benacerraf and Putnam 1964 and is reprinted in this volume.

1952. "Meaning Postulates." *Philosophical Studies*, vol. 3. Reprinted in Carnap 1947 (2nd ed.) and Feigl, Sellars, and Lehrer 1972.

Carroll, L. See under Dodgson, C. L.

Castañeda, H.-N. 1959. "Arithmetic and Reality." *Australasian Journal of Philosophy*, vol. 37. Reprinted in Benacerraf and Putnam 1964.

1961. "On Mathematical Proofs and Meaning." *Mind*, n.s. vol. 70.

Chihara, C. 1961. "Wittgenstein and Logical Compulsion." *Analysis*, vol. 21.

1963. "Mathematical Discovery and Concept Formation." *Philosophical Review*, vol. 72.

1973. *Ontology and the Vicious-Circle Principle.* Ithaca, N.Y.: Cornell University Press.

Church, A. 1932. "A Set of Postulates for the Foundation of Logic," part 1. *Annals of Mathematics*, 2nd ser. vol. 33.

1933. "A Set of Postulates for the Foundation of Logic," part 2. *Annals of Mathematics*, 2nd ser. vol. 34.

1934. "The Richard Paradox." *American Mathematical Monthly*, vol. 41.

1936–8. "A Bibliography of Symbolic Logic," parts 1 and 2. *Journal of Symbolic Logic*, vols. 1 and 3.

1944. *Introduction to Mathematical Logic, Part I. Annals of Mathematics Studies*, no. 13. Princeton: Princeton University Press; London: Oxford University Press, Milford. Revised and enlarged, *Princeton Mathematical Series*, no. 17 (1956).

1951. "Structure: A Formulation of the Logic of Sense and Denotation." In *Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer.* Edited by P. Henle, H. Kallen, and S. Langer. New York: Liberal Arts Press.

1956. "Propositions and Sentences." In I. M. Bocheński, A. Church, and N. Goodman, *The Problem of Universals: A Symposium* (Aquinas Symposium, University of Notre Dame, 1956). Notre Dame, Ind.: University of Notre Dame Press.

1958. "Ontological Commitment." *Journal of Philosophy*, vol. 55.

1974. "Set Theory with a Universal Set." In *Proceedings of the Tarski Symposium* (University of California, Berkeley, 1971). *Proceedings of Symposia in Pure Mathematics*, vol. 25. Edited by L. Henkin, J. Addison, C. C. Chang, W. Craig, D. Scott, and R. Vaught. Providence, R.I.: American Mathematical Society.

Chwistek, L. 1933. "Die nominalistische Grundlegung der Mathematik." *Erkenntnis*, vol. 3.

Cohen, M. R., and Nagel, E. 1934. *An Introduction to Logic and Scientific Method*. New York: Harcourt.

Cohen, P. J. 1963a. "The Independence of the Continuum Hypothesis," part 1. *Proceedings of the National Academy of Sciences*, vol. 50.

1963b. "A Minimal Model for Set Theory." *Bulletin of the American Mathematical Society*, vol. 69.

1964. "The Independence of the Continuum Hypothesis," part 2. *Proceedings of the National Academy of Sciences*, vol. 51.

1966. *Set Theory and the Continuum Hypothesis*. New York: Benjamin.

Curry, H. B. 1951. *Outlines of a Formalist Philosophy of Mathematics*. Amsterdam: North-Holland.

1954. "Remarks on the Definition and Nature of Mathematics." *Dialectica*, vol. 8. Reprinted in Benacerraf and Putnam 1964 and in this volume.

1958. "Calculuses and Formal Systems." *Dialectica*, vol. 12.

Davis, M., ed. 1965. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. Hewlett, N.Y.: Raven Press.

Dedekind, R. 1872. *Stetigkeit und irrationale Zahlen*. Brunswick: Vieweg. Reprinted in Dedekind 1930–2. Translated as "Continuity and Irrational Numbers" in Dedekind 1901.

1888. *Was sind und was sollen die Zahlen?* Brunswick: Vieweg. Reprinted in Dedekind 1920–2. Translated as "The Nature and Meaning of Numbers" in Dedekind 1901.

1901. *Essays on the Theory of Numbers*. Translations of Dedekind 1872 and 1888 by W. W. Beman. Chicago: Open Court; reprinted, New York: Dover, 1963.

1930–2. *Gesammelte mathematische Werke*, 3 vols. Edited by R. Fricke, E. Noether, and O. Ore. Brunswick: Vieweg.

Destouches, J.-L. 1951. "Sur la mécanique classique et l'intuitionisme." *Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen*, ser. A vol 54, and *Indagationes Mathematicae*, vol. 13.

Dieudonné, J. 1951. "L'axiomatique dans les mathématiques modernes." *XVII^e Congrès international de philosophie des sciences: Philosophie Mathématique/Mécanique* (Paris, 1949). *Actualités scientifiques et industrielles*, no. 1137. Paris: Hermann.

Dodgson, C. L. [Carroll, L.]. 1895. "What the Tortoise Said to Achilles." *Mind*, n.s. vol. 4.

Drake, F. R. 1974. *Set Theory: An Introduction to Large Cardinals*. Amsterdam: North-Holland.

Dreben, B., Andrews, P., and Aanderaa, S. 1963. "False Lemmas in Herbrand." *Bulletin of the American Mathematical Society*, vol. 69.

Dubislav, W. 1925. "Über das Verhältnis der Logik zur Mathematik." *Annalen der Philosophie*, vol. 5.

1930. "Über den sogenannten Gegenstand der Mathematik." *Erkenntnis*, vol. 1.

1932. *Die Philosophie der Mathematik in der Gegenwart. Philosophische Forschungsberichte*, vol. 13. Berlin: Junker and Dünnhaupt.

Du Bois-Reymond, P. 1869. "Bemerkungen über die verschiedenen Werthe, welche eine Function zweier reellen Variabeln erhält, wenn man diese Variabeln entweder nach einander oder gewissen Beziehungen gemäss gleichzeitig verschwinden lässt." *Journal für die reine und angewandte Mathematik*, vol. 70.

1873. "Eine neue Theorie der Convergenz und Divergenz von Reihen mit positiven Gliedern." *Journal für die reine und angewandte Mathematik*, vol. 76.

1875. "Ueber asymptotische Werte, infinitäre Approximationen und infinitäre Auflösung von Gleichungen." *Mathematische Annalen*, vol. 8.

1877. "Ueber die Paradoxen des Infinitärcalcüls." *Mathematische Annalen*, vol. 11.

1880. "Der Beweis des Fundamentalsatzes der Integralrechnung: $\int_a^b F'(x)\,dx = F(b) - F(a)$." *Mathematische Annalen*, vol. 16.

Dummett, M. 1955–6. "Frege on Functions: A Reply" and "Note: Frege on Functions." *Philosophical Review*, vols. 64 and 65. Reprinted as "Frege on Functions" in Dummett 1978.

1956. "Nominalism." *Philosophical Review*, vol. 65. Reprinted in Dummett 1978.

1957. "Constructionalism." *Philosophical Review*, vol. 66. Reprinted in Dummett 1978.

1959a. "Truth." *Proceedings of the Aristotelian Society* (1958–9), n.s. vol. 59. Reprinted in Dummett 1978.

1959b. "Wittgenstein's Philosophy of Mathematics." *Philosophical Review*, vol. 68. Reprinted in Benacerraf and Putnam 1964 and Dummett 1978.

1973. *Frege: Philosophy of Language*. London: Duckworth; New York: Harper and Row.

1975. "The Philosophical Basis of Intuitionistic Logic." In *Logic Colloquium '73. Proceedings of the Logic Colloquium* (Bristol, 1973). *Studies of the Logic and the Foundations of Mathematics*, vol. 80. Edited by H. E. Rose and J. C. Shepherdson. Amsterdam: North-Holland; New York: American Elsevier. Reprinted in Dummett 1978 and in this volume.

1977. *Elements of Intuitionism.* Oxford: Clarendon Press.
1978. *Truth and Other Enigmas.* Cambridge: Harvard University Press.
Erdmann, B. 1877. *Die Axiome der Geometrie.* Leipzig: Voss.
Erdmann, J. E., ed. 1840. *God. Guil. Leibnitii Opera Philosophica.* Berlin: Eichler.
Errera, A. 1953. "Le Problème du continu." *Atti della Accademia Ligure di Scienze e Lettere* (Genova), vol. 9.
Evans, G. 1973. "The Causal Theory of Names I." *Aristotelian Society Supplementary Volume 47.* Reprinted in Schwartz 1977.
Feferman, S. 1960. "Arithmetization of Metamathematics in a General Setting." *Fundamenta Mathematicae*, vol. 49.
Feigl, H. 1950. "Existential Hypotheses." *Philosophy of Science*, vol. 17.
Feigl, H., and Brodbeck, M., eds. 1953. *Readings in the Philosophy of Science.* New York: Appleton-Century-Crofts.
Feigl, H., and Sellars, W., eds. 1949. *Readings in Philosophical Analysis.* New York: Appleton-Century-Crofts.
Feigl, H., Sellars, W., and Lehrer, K., eds. 1972. *New Readings in Philosophical Analysis.* New York: Appleton-Century-Crofts.
Field, H. H. 1972. "Tarski's Theory of Truth." *Journal of Philosophy*, vol. 69.
1980. *Science without Numbers.* Princeton: Princeton University Press.
Findlay, J. N. 1942. "Gödelian Sentences: A Non-numerical Approach." *Mind*, n.s. vol. 51.
Fraenkel, A. A. 1919. *Einleitung in die Mengenlehre.* Berlin: Springer; 2nd ed., 1923; 3rd ed., 1928; reprinted, New York: Dover, 1946.
1930. "Die heutigen Gegensätze in der Grundlegung der Mathematik." *Erkenntnis*, vol. 1.
1935. "Sur la notion d'existence dans les mathématiques." *L'Enseignement mathématique*, 1st ser. vol. 34.
1953. *Abstract Set Theory.* Amsterdam: North-Holland; 2nd rev. ed., 1961; 4th rev. ed. by A. Lévy, North-Holland and New York: American Elsevier, 1976.
Fraenkel, A. A., and Bar-Hillel, Y. 1958. *Foundations of Set Theory.* Amsterdam: North-Holland; 2nd rev. ed. by A. A. Fraenkel, Y. Bar-Hillel, and A. Lévy, *Studies in Logic and the Foundations of Mathematics*, vol. 67. Amsterdam: North-Holland, 1973.
Frege, G. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Halle: Nebert. Reprinted in Frege 1964b. Translated as "Begriffsschrift, a Formula Language, Modeled upon that of Arithmetic, for Pure Thought" by S. Bauer-Mengelberg in van Heijenoort 1967 and as Frege 1972. Chapter 1 is translated by P. T. Geach in Frege 1952.
1884. *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl.* Breslau: Koebner. Translated as Frege 1950; excerpted and translated as Frege 1964c.
1885. "Über formale Theorien der Arithmetik." *Sitzungsberichte der Jenaischen Gesellschaft für Medicin und Naturwissenschaft für das Jahr 1885.* Jena: Fischer. Reprinted in Frege 1967. Translated in Frege 1971.

1981. *Function und Begriff. Vortrag gehalten in der Sitzung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medicin und Naturwissenschaft* (Jena). Reprinted in Frege 1962. Translated as "Function and Concept" by P. T. Geach in Frege 1952.
1892a. "Über Begriff und Gegenstand." *Vierteljahrsschrift für wissenschaftliche Philosophie*, vol. 16. Reprinted in Frege 1962 and 1967. Translated as "On Concept and Object" by P. T. Geach in Frege 1952 and reprinted in Feigl, Sellars, and Lehrer 1972.
1892b. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und philosophische Kritik*, n.s. vol. 100. Reprinted in Frege 1962. Translated as "On Sense and Nominatum" by H. Feigl in Feigl and Sellars 1949 and as "On Sense and Reference" by M. Black in Frege 1952.
1893–1903. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, 2 vols. Jena: Pohle; reprinted, Hildesheim: Olms, 1962. Partial translations in Frege 1952 and 1964a.
1897. "Über die Begriffsschrift des Herrn Peano und meine eigene." *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematische-Physische Klasse*, vol. 48. Translated as Frege 1969.
1903. "Über die Grundlagen der Geometrie." *Jahresberichte der Deutschen Mathematiker-Vereinigung*, vol. 12. Translated in Frege 1971.
1904. "Was ist eine Funktion?" *Festschrift Ludwig Boltzmann gewidmet zum sechzigsten Geburtstage 20 Februar 1904.* Leipzig: Barth. Translated as "What Is a Function?" by P. T. Geach in Frege 1952.
1919. "Der Gedanke – Die Verneinung: Zwei logische Untersuchungen." *Beiträge zur Philosophie des Deutschen Idealismus*, vol. 1. "Der Gedanke" is translated as Frege 1956. "Die Verneinung" is translated as "Negation" by P. T. Geach in Frege 1952. Both "Der Gedanke" ("The Thought") and "Die Verneinung" ("Negation") are translated in Frege 1977.
1950. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Numbers.* A translation, with German text, of Frege 1884 by J. L. Austin. Oxford: Blackwell; New York: Philosophical Library; 2nd rev. ed., 1953; reprinted, without German text, New York: Harper Torchbooks, 1960.
1952. *Translations from the Philosophical Writings of Gottlob Frege.* Edited and translated by P. T. Geach and M. Black. Oxford: Blackwell; 2nd ed., 1960. Includes Frege 1891, 1892a, 1892b, 1904, 1919, and selections from 1897 and 1893–1903.
1956. "The Thought: A Logical Inquiry." A translation of part 1 of Frege 1919 by A. Quinton and M. Quinton. *Mind*, n.s. vol. 65.
1962. *Funktion, Begriff, Bedeutung: Fünf logische Studien.* Edited by G. Patzig. Göttingen: Vandenhoeck and Ruprecht; 2nd rev. ed., 1965. Includes Frege 1891, 1892a, and 1892b.
1964a. *The Basic Laws of Arithmetic: Exposition of the System.* Translated and edited, with an introduction, by M. Furth. Berkeley and Los Angeles: University of California Press. Includes a partial translation of Frege 1893–1903.

1964b. *Begriffsschrift und andere Aufsätze*. Edited by I. Angelelli. Hildesheim: Olms. Includes Frege 1879.

1964c. "The Concept of Number." Excerpted from Frege 1884, and translated by M. S. Mahoney in Benacerraf and Putnam 1964 and reprinted in this volume.

1967. *Kleine Schriften*. Edited by I. Angelelli. Darmstadt and Hildesheim: Olms. Includes Frege 1885 and 1892a.

1969. "On Herr Peano's Begriffsschrift and My Own." A translation of Frege 1897 by V. H. Dudman. *Australasian Journal of Philosophy*, vol. 47.

1969–76. *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel*, 2 vols. Vol. 1 edited by H. Hermes, F. Kambartel, and F. Kaulbach; vol. 2 edited by G. Gabriel, H. Hermes, K. Kambartel, C. Thiel, and A. Veraart. Hamburg: Meiner. Vol. 1 is translated as Frege 1979; vol. 2 is translated as Frege 1980.

1971. *On the Foundations of Geometry and Formal Theories of Arithmetic*. Translations of Frege 1885 and 1903 by E.-H. Kluge. London and New Haven: Yale University Press.

1972. *Conceptual Notation and Related Articles*. A translation of Frege 1879 by T. W. Bynum. Oxford: Clarendon Press.

1977. *Logical Investigations*. Translated by P. T. Geach and R. H. Stoothoff. Oxford: Blackwell; New York: Yale University Press. Includes a translation of Frege 1919.

1979. *Gottlob Frege: Posthumous Writings*. A translation of Fege 1969-76, vol. 1, by P. Long, R. White, and R. Hargreaves. London: Blackwell; Chicago: University of Chicago Press.

1980. *Gottlob Frege: Philosophical and Mathematical Correspondence*. A translation and abridgement of Frege 1969–76, vol. 2. Translated by H. Kaal; abridged by B. McGuinness. London: Blackwell; Chicago: University of Chicago Press.

Freudenthal, H. 1936. "Zur intuitionistischen Deutung logischer Formeln." *Compositio Mathematica*, vol. 4.

Friedman, H. 1973. "The Consistency of Classical Set Theory Relative to a Set Theory with Intuitionistic Logic." *Journal of Symbolic Logic*, vol. 38.

Furstenberg, H. 1977. "Ergodic Behavior of Diagonal Measures and a Theorem of Szemerédi on Arithmetic Progressions." *Journal d'analyse mathématique*, vol. 31.

Gallen, D. 1975. *Intensional and Higher-Order Modal Logic*. Amsterdam: North-Holland.

Gasking, D. 1940. "Mathematics and the World." *Australasian Journal of Psychology and Philosophy*, vol. 18. Reprinted in Benacerraf and Putnam 1964.

Geach, P. T. 1951a. "Frege's *Grundlagen*." *Philosophical Review*, vol. 60.

1951b. "Symposium: On What There Is I." In *Freedom, Language, and Reality. Aristotelian Society Supplementary Volume 25*.

1956. "Discussion Note: On Frege's Way Out." *Mind*, n.s. vol. 65.

1959. "Russell on Meaning and Denoting." *Analysis*, vol. 19.

1961. "Frege." In G. E. M. Anscombe and P. T. Geach, *Three Philosophers*. Oxford: Blackwell; Ithaca, N.Y.: Cornell University Press.

Gentzen, G. 1938a. "Die gegenwärtige Lage in der mathematischen Grundlagenforschung." *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften*, n.s. vol. 4. Leipzig: Hirzel.

1938b. "Neue Fassung des Widerspruchsfreiheitbeweises für die reine Zahlentheorie." *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften*, n.s. vol. 4. Leipzig: Hirzel.

1969. *The Collected Papers of Gerhard Gentzen*. Edited by M. E. Szabo. Amsterdam: North-Holland.

Gergonne, J. D. 1819. "Essai sur la théorie des définitions." *Annales des mathématiques pures et appliquées*, vol. 9.

Gödel, K. 1931. "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I." *Monatshefte für Mathematik und Physik*, vol. 38. Translated as Gödel 1962 and as "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems" by J. van Heijenoort in van Heijenoort 1967.

1931–2a. "Eine Interpretation des intuitionistischen Aussagenkalküls." *Ergebnisse eines mathematischen Kolloquiums*, vol. 4.

1931–2b. "Zur intuitionistischen Arithmetik und Zahlentheorie." *Ergebnisse eines mathematischen Kolloquiums*, vol. 4.

1931–2c. "Zum intuitionistischen Aussagenkalkül." *Ergebnisse eines mathematischen Kolloquiums*, vol. 4.

1939. "Consistency-Proof for the Generalized Continuum-Hypothesis." *Proceedings of the National Academy of Sciences*, vol. 25.

1940. *The Consistency of the Axiom of Choice and of the Generalized Continuum-hypothesis with the Axioms of Set Theory*. *Annals of Mathematics Studies*, no. 3. Princeton: Princeton University Press; London: Oxford University Press, Milford; reprinted, Princeton University Press, 1951, 1953, and 1958.

1944. "Russell's Mathematical Logic." In *The Philosophy of Bertrand Russell*. Edited by P. A. Schilpp. Evanston and Chicago: Northwestern University; reprinted several times. Reprinted in Benacerraf and Putnam 1964 and in this volume.

1947. "What Is Cantor's Continuum Problem?" *American Mathematical Monthly*, vol. 54. Revised and expanded as Gödel 1964 in Benacerraf and Putnam 1964 and reprinted in this volume.

1958. "Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes." *Dialectica*, vol. 12.

1962. *On Formally Undecidable Propositions of* Principia Mathematica *and Related Systems*. A translation of Gödel 1931 by B. Meltzer, with an introduction by R. B. Braithwaite. Edinburgh: Oliver and Boyd.

1964. "What Is Cantor's Continuum Problem?" In Benacerraf and Putnam (1964) and reprinted in this volume.

Gogol, D. 1978. "The $\forall_n \exists$-Completeness of Zermelo-Fraenkel Set Theory."

Zeitschrift für Mathematische Logik und Grundlagen der Mathematik, vol. 24.

Goldman, A. I. 1967. "A Causal Theory of Knowing." *Journal of Philosophy*, vol. 64.

Goodman, N. 1951. *The Structure of Appearance*. Cambridge: Harvard University Press.

1956. "A World of Individuals." In I. M. Bocheński, A. Church, and N. Goodman, *The Problem of Universals: A Symposium* (Aquinas Symposium, University of Notre Dame, 1956). Notre Dame, Ind.: University of Notre Dame Press. Reprinted in Benacerraf and Putnam 1964.

Goodman, N., and Quine, W. V. 1947. "Steps Toward a Constructive Nominalism." *Journal of Symbolic Logic*, vol. 12.

Goodstein, R. L. 1957. "Critical Notice: Wittgenstein's *Remarks on the Foundations of Mathematics*." *Mind*, n.s. vol. 66.

1958. "On the Nature of Mathematical Systems." *Dialectica*, vol. 12.

1962. "Symposium: The Foundations of Mathematics I." *Aristotelian Society Supplementary Volume 36*.

Grelling, K. 1936. "The Logical Paradoxes." *Mind*, n.s. vol. 45.

Grice, H. P., and Strawson, P. F. 1956. "In Defense of a Dogma." *Philosophical Review*, vol. 65. Reprinted in Feigl, Sellars, and Lehrer 1972.

Griss, G. F. C. 1946a. *Idealistische Filosofie: een humanistische levens- en wereldbeschouwing*. Arnhem: Van Loghum Slaterus.

1946b. "Negationless Intuitionistic Mathematics." *Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen*, vol. 49, and *Indagationes Mathematicae*, vol. 8.

Haack, S. 1974. *Deviant Logics: Some Philosophical Issues*. Cambridge University Press.

Hahn, H. 1933. *Logik, Mathematik und Naturerkennen. Einheitswissenschaft*, vol. 2. Vienna: Gerold. Translated as "Logic, Mathematics and Knowledge of Nature" by A. Pap in Ayer 1959.

Hajnal, A. 1956. "On a Consistency Theorem Connected with the Generalized Continuum Problem." *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 2.

Hanf, W. P., and Scott, D. 1961. "Classifying Inaccessible Cardinals." *Notices of the American Mathematical Society*, vol. 8.

Hardy, G. H. 1929. "Mathematical Proof." *Mind*, n.s. vol. 38.

Harman, G. H. 1973. *Thought*. Princeton: Princeton University Press.

Hart, W. P. 1970. "Skolem's Promises and Paradoxes." *Journal of Philosophy*, vol. 67.

1974. "On an Argument for Formalism." *Journal of Philosophy*, vol. 71.

Hausdorff, F. 1914. *Grundzüge der Mengenlehre*. Leipzig: Veit; 2nd rev. ed., Berlin and Leipzig: de Gruyter, 1927; 3rd ed., 1935. Translated as Hausdorff 1949.

1949. *Set Theory*. Translated from the 3rd German edition of Hausdorff 1914 by J. R. Auman. New York: Chelsea; reprinted, 1957, 1962.

Hempel, C. G. 1945a. "Geometry and Empirical Science." *American Mathematical Monthly*, vol. 52. Reprinted in Feigl and Sellars 1949.

1945b. "On the Nature of Mathematical Truth." *American Mathematical Monthly*, vol. 52. Reprinted in Feigl and Sellars 1949, Feigl and Brodbeck 1953, Benacerraf and Putnam 1964, and in this volume.

Hempel, C. G., and Oppenheim, P. 1936. *Der Typusbegriff im Lichte der neuen Logik*. Leiden: Sijthoff.

Henkin, L. 1949. "The Completeness of the First-Order Functional Calculus." *Journal of Symbolic Logic*, vol. 14. Reprinted in Hintikka 1969.

1950. "Completeness in the Theory of Types." *Journal of Symbolic Logic*, vol. 15. Reprinted in Hintikka 1969.

Herbrand, J. 1968. *Écrits logiques*. Edited by J. van Heijenoort. Paris: Presses Universitaires de France. Translated as Herbrand 1972.

1972. *Logical Writings*. A translation of Herbrand 1968. Edited by W. D. Goldfarb. Dordrecht: Reidel; Cambridge: Harvard University Press.

Heyting, A. 1931. "Die intuitionistische Grundlegung der Mathematik." *Erkenntnis*, vol. 2. Translated as "The Intuitionist Foundations of Mathematics" by E. Putnam and G. J. Massey in Benacerraf and Putnam 1964 and reprinted in this volume.

1934. *Mathematische Grundlagenforschung, Intuitionismus, Beweistheorie*. Berlin: Springer.

1948. "Formal Logic and Mathematics." *Synthèse*, vol. 6.

1956. *Intuitionism: an Introduction*. Amsterdam: North-Holland; 2nd rev. ed., 1966; 3rd rev. ed., 1971. Excerpted as "Disputation" in Benacerraf and Putnam 1964 and reprinted in this volume.

1958. "Blick von der intuitionistischen Warte." *Dialectica*, vol. 12.

1961. "Axiomatic Method and Intuitionism." In *Essays on the Foundations of Mathematics, Dedicated to A. A. Fraenkel on His Seventieth Anniversary*. Edited by Y. Bar-Hillel, E. I. J. Poznanski, M. O. Rabin, and A. Robinson. Jerusalem: Magnes Press, Hebrew University.

Hilbert, D. 1899a. "Grundlagen der Geometrie." *Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen*. Leipzig: Teubner. Reprinted as Hilbert 1899b.

1899b. *Grundlagen der Geometrie*. Leipzig: Teubner; 2nd ed., 1903; 3rd rev. ed., 1909; 7th rev. and enlarged ed., 1930; 10th ed., with a supplement by P. Bernays, Stuttgart, 1968. Translated as Hilbert 1902 and 1971.

1902. *The Foundations of Geometry*. A translation of Hilbert 1899b by E. J. Townsend. Chicago: Open Court.

1905a. "Über die Grundlagen der Logik und der Arithmetik." *Verhandlungen des Dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*. Lepizig: Teubner. Reprinted in Hilbert 1899b (3rd ed.). Translated as Hilbert 1905b and as "On the Foundations of Logic and Arithmetic" by B. Woodward in van Heijenoort 1967.

1905b. "On the Foundations of Logic and Arithmetic." A translation of Hilbert 1905a by G. B. Halsted. *Monist*, vol. 15.

1918. "Axiomatisches Denken." *Mathematische Annalen*, vol. 78. Reprinted in Hilbert 1932–5, vol. 3.

1922. "Neubegründung der Mathematik." *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, vol. 1. Reprinted in Hilbert 1932–5, vol. 3.

1926. "Über das Unendliche." *Mathematische Annalen*, vol. 95. Translated as "On the Infinite" by E. Putnam and G. J. Massey in Benacerraf and Putnam 1964 and by S. Bauer-Mengelberg in van Heijenoort 1967. Reprinted in this volume.

1928. "Die Grundlagen der Mathematik." *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, vol. 6. Reprinted in Hilbert 1899 (7th ed.) and in Hilbert, Weyl, and Bernays 1928. Translated as "The Foundations of Mathematics" by S. Bauer-Mengelberg and D. Føllesdal in van Heijenoort 1967.

1931. "Die Grundlegung der elementaren Zahlenlehre." *Mathematische Annalen*, vol. 104.

1932–5. *Gesammelte Abhandlungen*, 3 vols. Berlin: Springer; 2nd ed., 1970; 3rd ed., New York: Chelsea, 1981.

1971. *The Foundations of Geometry*. A translation of Hilbert 1899b (10th ed.) by L. Under. La Salle, Ill.: Open Court.

Hilbert, D., and Bernays, P. 1934–9. *Grundlagen der Mathematik*, 2 vols. Berlin: Springer; 2nd ed., 1968–70.

Hilbert, D., Weyl, H., and Bernays, P. 1928. *Die Grundlagen der Mathematik. Hamburger Mathematische Einzelschriften*, vol. 5. Leipzig: Teubner. Includes Hilbert 1928.

Hintikka, K. J. J. 1959. "Existential Presuppositions and Existential Commitments." *Journal of Philosophy*, vol. 56.

1969a. "On Kant's Concept of Intuition (*Anschauung*)." In *The First Critique: Reflections on Kant's Critique of Pure Reason*. Edited by T. Penelhum and J. J. MacIntosh. Belmont, Calif.: Wadsworth.

1969b. Ed. *The Philosophy of Mathematics*. Oxford: Oxford University Press.

1973. *Logic, Language-Games, and Information*. Oxford: Clarendon Press.

1974. *Knowledge and the Known: Historical Perspectives in Epistemology*. Dordrecht: Reidel.

Hook, S., ed. 1956. *American Philosophers at Work*. New York: Criterion.

Huntington, E. V. 1913. "A Set of Postulates for Abstract Geometry." *Mathematische Annalen*, vol. 73.

Hurewicz, W. 1932. "Une remarque sur l'hypothèse du continu." *Fundamenta Mathematicae*, vol. 19.

Jockusch, C. G., Jr. 1972. "Ramsey's Theorem and Recursion Theory." *Journal of Symbolic Logic*, vol. 37.

Jourdain, P. E. B. 1906–13. "The Development of the Theory of Transfinite Numbers," parts 1–4. *Archiv der Mathematik und Physik*, vols. 10, 14, 16, 22.

Jubien, M. 1977. "Ontology and Mathematical Truth." *Noûs*, vol. 11.

Kamke, E. 1928. *Mengenlehre*. Berlin and Leipzig: Gruyter; 2nd ed., 1947. Translated as Kamke 1950.

1950. *Theory of Sets*. A translation of Kamke 1928 (2nd ed.) by F. Bagemihl. New York: Dover.

Kant, I. 1781. *Kritik der reinen Vernunft*. Riga; 2nd ed., 1787. Translated as Kant 1881.

1800. *Immanuel Kants Logik, ein Handbuch zu Vorlesungen*. Edited by G. B. Jäsche. Königsberg: Nicolovius. Reprinted in Kant 1867–8, vol. 8.

1867–8. *Immanuel Kants Sämmtliche Werke*, 8 vols. Edited by G. Hartenstein. Leipzig: Voss.

1881. *Critique of Pure Reason*. A translation of Kant 1781 (2nd ed.) by F. M. Müller. London: Macmillan; 2nd rev. ed., 1896; 5th printing, 1953.

Kaufmann, F. 1930. *Das Unendliche in der Mathematik und seine Ausschaltung*. Leipzig and Vienna: Deuticke.

1931. "Bemerkungen zum Grundlagenstreit in Logik und Mathematik." *Erkenntnis*, vol. 2.

Kazemier, B. H., and Vuysje, D., eds. 1962. *Logic and Language: Studies Dedicated to Professor Rudolf Carnap on the Occasion of His Seventieth Birthday*. Dordrecht: Reidel.

Kino, A., Myhill, J., and Vesley, R., eds. 1970. *Intuitionism and Proof Theory. Proceedings of the Summer Conference at Buffalo, N.Y.*, 1968. Amsterdam: North-Holland.

Kleene, S. C. 1949. "On the Intuitionistic Logic." In *Proceedings of the Tenth International Congress of Philosophy* (Amsterdam, 1948), vol. 1, pt. 2. Edited by E. W. Beth, H. J. Pos, and J. H. A. Hollak. Amsterdam: North-Holland.

1952. *Introduction to Metamathematics*. Amsterdam: North-Holland; Groningen: Noordhoff; New York: Van Nostrand.

Kleene, S. C., and Rosser, J. B. 1935. "The Inconsistency of Certain Formal Logics." *Annals of Mathematis*, vol. 36.

Kneale, W. 1946. "The Truths of Logic." *Proceedings of the Aristotelian Society* (1945–6), n.s. vol. 46.

Kneale, W., and Kneale, M. 1962. *The Development of Logic*. Oxford: Clarendon Press.

König, J. 1905. "Zum Kontinuum-Problem." *Mathematische Annalen*, vol. 60.

Körner, S. 1960. *The Philosophy of Mathematics: An Introduction*. London: Hutchinson; reprinted, New York: Harper and Brothers, 1962.

Kreisel, G. 1950. "Note on Arithmetic Models for Consistent Formulae of the Predicate Calculus." *Fundamenta Mathematicae*, vol. 37.

1951. "Some Remarks on the Foundations of Mathematics: An Expository Article." *Mathematical Gazette*, vol. 35.

1951–2. "On the Interpretation of Non-finitist Proofs," parts 1 and 2. *Journal of Symbolic Logic*, vols. 16 and 17.

1953a. "The Diagonal Method in Formalized Arithmetic." *British Journal for the Philosophy of Science*, vol. 3.

1953b. "On a Problem of Henkin's." *Indagationes Mathematicae*, vol. 15.

1953c. "A Variant to Hilbert's Theory of the Foundations of Arithmetic." *British Journal for the Philosophy of Science*, vol. 4.

1958a. "Hilbert's Programme." *Dialectica*, vol. 12. An expansion of the revised version published in Benacerraf and Putnam 1964 is reprinted in this volume.
1958b. "Mathematical Significance of Consistency Proofs." *Journal of Symbolic Logic*, vol. 23.
1958c. "Relative Consistency Proofs." *Journal of Symbolic Logic*, vol. 23.
1958d. "A Remark on Free Choice Sequences and the Topological Completeness Proofs." *Journal of Symbolic Logic*, vol. 23.
1959. "Interpretation of Analysis by Means of Constructive Functionals of Finite Types." In *Constructivity in Mathematics: Proceedings of the Colloquium* (Amsterdam, 1957). Edited by A. Heyting. Amsterdam: North-Holland.
1960a. "Ordinal Logics and the Characterization of Informal Concepts of Proof." *Proceedings of the International Congress of Mathematicians, 1958* (Edinburgh). Cambridge University Press.
1960b. "La Prédicativité." *Bulletin de la Société Mathématique de France*, vol. 88.
1962a. "Foundations of Intuitionistic Logic." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (Stanford, Calif.). Edited by E. Nagel, P. Suppes, and A. Tarski. Stanford, Calif.: Stanford University Press.
1962b. Review of "Arithmetization of Metamathematics in a General Setting" (Feferman 1960). *Zentralblatt für Mathematik und ihre Grenzgebiete*, vol. 95.
1962c. Review of *Beweistheorie* (Schütte 1960). *Mathematical Gazette*, vol. 46.
1965. "Mathematical Logic." In *Lectures on Modern Mathematics*, vol. 3. Edited by T. L. Saaty. New York: Wiley.
1967. "Informal Rigour and Completeness Proofs." In *Problems in the Philosophy of Mathematics*. Edited by I. Lakatos. Amsterdam: North-Holland.
1968. "A Survey of Proof Theory." *Journal of Symbolic Logic*, vol. 33.
1976a. "Relative Consistency Proofs. II." *Journal of Symbolic Logic*, vol. 41.
1976b. "What Have We Learned from Hilbert's Second Problem?" In *Mathematical Developments Arising from Hilbert Problems. Proceedings of Symposia in Pure Mathematics* (Northern Illinois University, DeKalb, 1974), vol. 28. Edited by F. E. Browder. Providence, R.I.: American Mathematical Society.
1976c. "Wie die Beweistheorie zu ihren Ordinalzahlen kam und kommt." *Jahresberichte der Deutschen Mathematiker-Vereinigung*, vol. 78.
1980. Obituary of Kurt Gödel. *Biographical Memoirs of Fellows of the Royal Society*, vol. 28, with corrigenda and addenda in vols. 29 (1981) and 30 (1982).
1982. "Finiteness Theorems in Arithmetic: An Application of Herbrand's Theorem for Σ_2-Formulas." *Proceedings of the Herbrand Symposium, Logic Colloquium 1981*. Edited by J. Stern. Amsterdam: North-Holland.
Kreisel, G., and Takeuti, G. 1974. "Formally Self-referential Propositions for Cut-Free Classical Analysis and Related Systems." *Dissertationes Mathematicae*, vol. 118.
Kreisel, G., Mints, G. E., and Simpson, S. G. 1975. "The Use of Abstract Language in Elementary Metamathematics: Some Pedagogic Examples." In

Logic Colloquium (Boston, 1972–3). Edited by R. Parikh. *Lecture Notes in Mathematics*, no. 453. Edited by A. Dold and B. Eckmann. Berlin, Heidelberg, and New York: Springer.
Kripke, S. 1965. "Semantic Analysis of Intuitionistic Logic I." In *Formal Systems and Recursive Functions: Proceedings of the Eighth Logic Colloquium* (Oxford, 1963). Edited by J. N. Crossley and M. A. E. Dummett. Amsterdam: North-Holland.
1972. "Naming and Necessity." In *Semantics of Natural Language*. Edited by D. Davidson and G. Harman. Dordrecht: Reidel.
1975. "Outline of a Theory of Truth." *Journal of Philosophy*, vol. 72.
Kuratowski, C. 1933–50. *Topologie*, 2 vols. *Monografie Matematyczne*, vols. 3 and 21; 2nd ed., Warsaw: Państwowe Wydawnictwo Naukowe, 1948–52. Translated as Kuratowski 1966–8.
1948. "Ensembles projectifs et ensembles singuliers." *Fundamenta Mathematicae*, vol. 35.
1951. "Sur une charactérisation des alephs." *Fundamenta Mathematicae*, vol. 38.
1966–8. *Topology*, 2 vols. A translation of Kuratowski 1933–50 by J. Jaworowski. Warsaw: Pańtwowe Wydawnictwo Naukowe; London and New York: Academic Press.
Lakatos, I. 1962. "Symposium: The Foundations of Mathematics II." *Aristotelian Society Supplementary Volume 36*.
1967. Ed. *Problems in the Philosophy of Mathematics*. Amsterdam: North-Holland.
Langford, C. H. 1927. "On Inductive Relations." *Bulletin of the American Mathematical Society*, vol. 33.
Lear, J. 1977. "Sets and Semantics." *Journal of Philosophy*, vol. 74.
Leibniz, G. W. von. 1875–90. *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz*, 7 vols. Edited by C. J. Gerhardt. Berlin: Weidmann; facsimile reprint, Hildesheim: Olms, 1962.
1923- *Sämtliche Schriften und Briefe*. Edited by the Preussische Akademie der Wissenschaften. Darmstadt: Reichl.
Lévy, A. 1960. "Axiom Schemata of Strong Infinity in Axiomatic Set Theory." *Pacific Journal of Mathematics*, vol. 10.
Lewis, C. I. 1918. *A Survey of Mathematical Logic*. Berkeley: University of California Press.
Lewis, C. I., and Langford, C. H. 1932. *Symbolic Logic*. New York and London: Century; 2nd ed., New York: Dover, 1959.
Lewis, D. K. 1968. "Counterpart Theory and Quantified Modal Logic." *Journal of Philosophy*, vol. 65.
1969. *Convention*. Cambridge: Harvard University Press.
Lewy, C. 1946. "Symposium: Why Are the Calculuses of Logic and Mathematics Applicable to Reality?" part 2. In *Logic and Reality. Aristotelian Society Supplementary Volume 20*.
Linsky, L. 1952. *Semantics and the Philosophy of Language*. Urbana: University of Illinois Press.

Lorenzen, P. 1955. *Einführung in die operative Logik und Mathematik. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen*, vol. 78. Berlin: Springer.

1957. "Wie ist Philosophie der Mathematik möglich?" *Philosophia Naturalis*, vol. 4.

1962. *Metamathematik*. Mannheim: Bibliographisches Institut.

Łoś, J., Mostowski, A., and Rasiowa, H. 1956. "A Proof of Herbrand's Theorem." *Journal des mathématiques pures et appliquées*, vol. 35; correction, vol. 40 (1961).

Lucas, J. R. 1961. "Minds, Machines, and Gödel." *Philosophy: The Journal of the Royal Institute of Philosophy*, vol. 36.

Łukasiewicz, J. 1929. *Elementy logiki matematycznej*. Authorized lecture notes prepared by M. Presburger. Warsaw: Warsaw University; 2nd ed., Warsaw: Państwowe Wydawnictwo Naukowe, 1958. Translated as Łukasiewicz 1963.

1941. "Die Logik und das Grundlagenproblem." *Les Entretiens de Zurich sur les fondements et la méthode des sciences mathématiques* (1938). Edited by F. Gonseth. Zürich: Leemann.

1963. *Elements of Mathematical Logic*. A translation of Łukasiewicz 1929 by O. Wojtasiewicz. Oxford: Pergamon Press; New York: Macmillan.

Lusin, N. 1914. "Sur un problème de M. Baire." *Comptes Rendus des Séances de l'Académie des Sciences* (Paris), vol. 158.

1935. "Sur les ensembles analytiques nuls." *Fundamenta mathematicae*, vol. 25.

Lusin, N., and Sierpiński, W. 1918. "Sur quelques propriétés des ensembles." *Bulletin International de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques et Naturelles, ser. A: Sciences Mathématiques*.

McCall, S., ed. 1967. *Polish Logic 1920–1939*. Oxford: Clarendon Press.

MacLane, S. 1961. "Locally Small Categories and the Foundations of Set Theory." In *Infinitistic Methods. Proceedings of the Symposium on Foundations of Mathematics* (Warsaw, 1959). London and New York: Pergamon Press.

McNaughton, R. 1957. "Conceptual Schemes in Set Theory." *Philosophical Review*, vol. 66.

Mahlo, P. 1911. "Über lineare transfinite Mengen." *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften*, vol. 63. Leipzig: Teubner.

1913. "Zur Theorie und Anwendung der ρ_0 = Zahlen. II." *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften*, vol. 65. Leipzig: Teubner.

Mannoury, G. 1909. *Methodologisches und Philosophisches zur Elementar-Mathematik*. Haarlem: Visser; Assen: Van Gorcum.

Marcus, R. B. 1974. "Classes, Collections, and Individuals." *American Philosophical Quarterly*, vol. 11.

Marshall, W. 1953. "Frege's Theory of Functions and Objects." *Philosophical Review*, vol. 62.

Martin, D. A. 1968. "The Axiom of Determinateness and Reduction Principles in the Analytic Hierarchy." *Bulletin of the American Mathematical Society*, vol. 74.

1976. "Hilbert's First Problem: the Continuum Hypothesis." In *Mathematical Developments Arising from Hilbert Problems. Proceedings of Symposia in Pure Mathematics* (Northern Illinois University, DeKalb, 1974), vol. 28. Edited by F. E. Browder. Providence, R.I.: American Mathematical Society.

Martin, R. M. 1958. *Truth and Denotation*. Chicago: University of Chicago Press.

1963. *Intension and Decision: A Philosophical Study*. Englewood Cliffs, N.J.: Prentice-Hall.

Mehlberg, H. 1962. "The Present Situation in the Philosophy of Mathematics." In *Logic and Language: Studies Dedicated to Professor Rudolf Carnap on the Occasion of His Seventieth Birthday*. Edited by B. H. Kazemier and D. Vuysje. Dordrecht: Reidel.

Menger, K. 1930. "Der Intuitionismus." *Blätter für deutsche Philosophie*, vol. 4.

1933. "Die neue Logik." In *Krise und Neuaufbau in den exakten Wissenschaften*. Leipzig and Vienna: Deuticke.

Meyer, R. 1975. "Weak Monadic Second Order Theory of Successor Is Not Elementary-Recursive." In *Logic Colloquium* (Boston, 1972-3). Edited by R. Parikh. *Lecture Notes in Mathematics*, no. 453. Edited by A. Dold and B. Eckmann. Berlin, Heidelberg, and New York: Springer.

Mill, J. S. 1843. *A System of Logic*. London: Longmans.

Mirimanoff, D. 1917a. "Les Antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles." *L'Enseignement mathématique*, 1st ser. vol. 19.

1917b. "Remarques sur la théorie des ensembles et les antinomies cantoriennes," part 1. *L'Enseignement mathématique*, 1st ser. vol. 19.

1920. "Remarques sur la théorie des ensembles et les antinomies cantoriennes," part 2. *L'Enseignement mathématique*, 1st ser. vol. 21.

Montague, R. 1970. "Pragmatics and Intensional Logic." *Dialectica*, vol. 24, and *Synthèse*, vol. 22. Reprinted in Montague 1974.

1974. *Formal Philosophy*. Edited, with an introduction, by R. H. Thomason. London and New Haven: Yale University Press.

Moore, G. E. 1954-5. "Wittgenstein's Lectures in 1930-33." *Mind*, n.s. vols. 63 and 64. Reprinted in Moore 1959.

1959. *Philosophical Papers*. London: Allen and Unwin; New York: Macmillan.

Moschovakis, Y. N. 1971. "Predicative Classes." In *Axiomatic Set Theory. Proceedings of Symposia in Pure Mathematics*, vol. 13, pt. 1. Edited by D. S. Scott. Providence, R.I.: American Mathematical Society.

Mostowski, A. 1957a. "On Recursive Models of Formalised Arithmetic." *Bulletin de L'Académie Polonaise des Sciences, Série des Sciences Mathématiques, Astronomiques et Physiques*, vol. 5.

1957b. Review of G. Kreisel, "A Variant to Hilbert's Theory of the Foundations of Arithmetic" (Kreisel 1953c). *Journal of Symbolic Logic*, vol. 22.

1965. "On Models of Zermelo-Fraenkel Set Theory Satisfying the Axiom of

Constructibility." In *Studia Logico-Mathematica et Philosophica. Acta Philosophica Fennica*, no. 18. Helsinki: Societas Philosophica Fennica.

1966. *Thirty Years of Foundational Studies. Acta Philosophica Fennica*, no. 17. New York: Barnes and Noble.

Mycielski, J. 1964. "On the Axiom of Determinateness." *Fundamenta Mathematicae*, vol. 53.

Myhill, J. 1951-2. "Two Ways of Ontology in Modern Logic." *Review of Metaphysics*, vol. 5.

1952-3. "Some Philosophical Implications of Mathematical Logic: Three Classes of Ideas." *Review of Metaphysics*, vol. 6.

Nagel, E. 1944. "Logic Without Ontology." In *Naturalism and the Human Spirit*. Edited by Y. H. Krikorian. New York: Columbia University Press. Reprinted in Feigl and Sellars 1949 and Benacerraf and Putnam 1964.

1948. "Book Review" of *Meaning and Necessity* (Carnap 1947). *Journal of Philosophy*, vol. 45.

1956. *Logic Without Metaphysics*. New York: Free Press.

Nagel, E., and Newman, J. R. 1958. *Gödel's Proof*. New York: New York University Press.

Natorp, P. G. 1910. *Die logischen Grundlagen der exakten Wissenschaften. Wissenschaft und Hypothese*, vol. 12. Lepizig and Berlin: Teubner.

Noteler, E., and Cavaillès, J., eds. 1937. *Briefwechsel Cantor-Dedekind. Actualités scientifiques et industrielles*, no. 518. Paris: Hermann.

Pap, A. 1958. *Semantics and Necessary Truth*. New Haven: Yale University Press.

1960. "Types and Meaninglessness." *Mind*, n.s. vol. 69.

Parikh, R., ed. 1975. *Logic Colloquium* (Boston, 1972-3). *Lecture Notes in Mathematics*, no. 453. Edited by A. Dold and B. Eckmann. Berlin, Heidelberg, and New York: Springer.

Paris, J., and Harrington, L. 1977. "A Mathematical Incompleteness in Peano Arithmetic." In *Handbook of Mathematical Logic*. Edited by J. Barwise. *Studies in Logic and the Foundations of Mathematics*, vol. 90. Amsterdam and New York: North-Holland.

Parsons, C. 1965. "Frege's Theory of Number." In *Philosophy in America*. Edited by M. Black. London: Allen and Unwin.

1969. "Kant's Philosophy of Arithmetic." In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. Edited by S. Morgenbesser, P. Suppes, and M. White. New York: St. Martin's Press.

1971. "Ontology and Mathematics." *Philosophical Review*, vol. 80.

1974a. "Informal Axiomatization, Formalization, and the Concept of Truth." *Synthèse*, vol. 27.

1974b. "The Liar Paradox." *Journal of Philosophical Logic*, vol. 3.

1974c. "Sets and Classes." *Noûs*, vol. 8.

1974d. "Sets and Possible Worlds." Abstract of a talk at Columbia University, mimeographed.

1976. "Some Remarks on Frege's Conception of Extension." In *Studien zu Frege = Studies on Frege*, 3 vols. Vol. 1, *Logik und Philosophie der Mathematik*. Edited by M. Schirn. Stuttgart: Frommann-Holzboog.

1977. "What Is the Iterative Conception of Set?" In *Logic, Foundations of Mathematics, and Computability Theory. Proceedings of the Fifth International Congress of Logic, Methodology and the Philosophy of Science* (London, Ontario, 1975), pt. 1. Edited by R. E. Butts and J. Hintikka. Dordrecht: Reidel. Reprinted in this volume.

1981. "Modal Set Theories." *Journal of Symbolic Logic*, vol. 46.

Forthcoming. "Quine on the Philosophy of Mathematics." In C. Parsons, *Mathematics in Philosophy*. Ithaca, N.Y.: Cornell University Press.

Peano, G. 1889. *Arithmetices principia, nova methodo exposita*. Turin: Bocca. Translated as "The Principles of Arithmetic, Presented by a New Method" by J. van Heijenoort in van Heijenoort 1967.

Peirce, C. S. 1931-58. *Collected Papers*, 8 vols. Edited by C. Hartshorne, P. Weiss, and A. W. Burks. Cambridge: Harvard University Press.

Plato. *Meno*. Translated by B. Jowett. New York: Liberal Arts Press; Indianapolis: Bobbs-Merrill, 1949.

Poincaré, H. 1894. "Sur la nature du raisonnement mathématique." *Revue de métaphysique et de morale*, vol. 2. Reprinted in Poincaré 1903. Translated as "On the Nature of Mathematical Reasoning" in Poincaré 1907. Excerpted and reprinted in Benacerraf and Putnam 1964 and in this volume.

1903. *La Science et l'Hypothèse*. Paris: Flammarion. Translated as Poincaré 1907 and in 1913b.

1905. "Les Mathématiques et la logique." *Revue de métaphysique et de morale*, vol. 13.

1907. *Science and Hypothesis*. A translation of Poincaré 1903 by "W. J. G." (W. J. Greenstreet). London: Scott; New York: Scribner; reprinted, New York: Dover, 1952.

1908a. "L'Avenir des mathématiques." *Scientia: Rivista di Scienza*, vol. 4, and *Revue générale des sciences pures et appliquées*, vol. 19. Reprinted in Poincaré 1908b. Translated in Poincaré 1913b and 1914.

1908b. *Science et Méthode*. Paris: Flammarion. Translated in Poincaré 1913b and as 1914.

1913a. *Dernières pensées. Bibliothèque de philosophie scientifique*. Paris: Flammarion.

1913b. *The Foundations of Science*. Translated by G. B. Halsted. New York: Science Press. Includes Poincaré 1903 and 1908b.

1914. *Science and Method*. A translation of Poincaré 1908b by F. Maitland. London and New York: Nelson; reprinted, New York: Dover, 1952.

Popper, K. 1946. "Symposium: Why Are the Calculuses of Logic and Mathematics Applicable to Reality?" part 3. In *Logic and Reality. Aristotelian Society Supplementary Volume 20*.

1947. "Logic Without Assumptions." *Proceedings of the Aristotelian Society* (1946-7), n.s. vol. 47.

1947-8. "New Foundations for Logic" and "Corrections and Additions to 'New Foundations for Logic.'" *Mind*, n.s. vols. 56 and 57.

1948. "What Can Logic Do for Philosophy?" In *Logical Positivism and Ethics. Aristotelian Society Supplementary Volume 22*.

Powell, W. C. 1972. "Set Theory with Predication." Ph.D. dissertation, State University of New York at Buffalo.

1975. "Extending Gödel's Negative Interpretation to ZF." *Journal of Symbolic Logic*, vol. 40.

Putnam, H. 1956. "Mathematics and the Existence of Abstract Entities." *Philosophical Studies*, vol. 7.

1962. "The Analytic and the Synthetic." In *Minnesota Studies in the Philosophy of Science*, vol. 1. Edited by H. Feigl and G. Maxwell. Minneapolis: University of Minnesota Press.

1967a. "Mathematics without Foundations." *Journal of Philosophy*, vol. 64. Reprinted in Putnam 1975, vol. 1, and in this volume.

1967b. "The Thesis that Mathematics Is Logic." In *Bertrand Russell, Philosopher of the Century*. Edited by R. Schoenman. London: Allen and Unwin.

1975. *Philosophical Papers*, 2 vols. Vol. 1: *Mathematics, Matter and Method;* vol. 2: *Mind, Language, and Reality*. Cambridge University Press.

1980. "Models and Reality." Presidential address to the annual meeting of the Association for Symbolic Logic, December 1977. *Journal of Symbolic Logic*, vol. 45. Reprinted in this volume.

Quine, W. V. 1936. "Truth by Convention." In *Philosophical Essays for Alfred North Whitehead*. Edited by O. H. Lee. New York: Longmans, Green. Reprinted in Feigl and Sellars 1949, Benacerraf and Putnam 1964, Quine 1966b, and in this volume.

1937. "New Foundations for Mathematical Logic." *American Mathematical Monthly*, vol. 44. Reprinted in Quine 1953a.

1939. "Designation and Existence." *Journal of Philosophy*, vol. 36. Reprinted in Feigl and Sellars 1949.

1940. *Mathematical Logic*. New York: Norton; rev. ed., Cambridge: Harvard University Press, 1951; reprinted, New York: Harper Torchbooks, 1962.

1941. "Whitehead and the Rise of Modern Logic." In *The Philosophy of Alfred North Whitehead*. Edited by P. A. Schilpp. Evanston and Chicago: Northwestern University Press; 2nd ed., New York: Tudor, 1951. Reprinted in Quine 1966a.

1943. "Notes on Existence and Necessity." *Journal of Philosophy*, vol. 40. Reprinted in Linsky 1952.

1947. "On Universals." *Journal of Symbolic Logic*, vol. 12.

1948. "On What There Is." *Review of Metaphysics*, vol. 2. Reprinted in "Symposium: On What There is III" in *Freedom, Language, and Reality. Aristotelian Society Supplementary Volume 25* (1951). Reprinted in Linsky 1952, Quine 1953a, Benacerraf and Putnam 1964, and Feigl, Sellars, and Lehrer 1972.

1951a. "On Carnap's Views on Ontology." *Philosophical Studies*, vol. 2. Reprinted in Quine 1966b and Feigl, Sellars, and Lehrer 1972.

1951b. "Ontology and Ideology." *Philosophical Studies*, vol. 2. Reprinted in Feigl, Sellars, and Lehrer 1972.

1951c. "Semantics and Abstract Objects." In *Contributions to the Analysis and Synthesis of Knowledge. Proceedings of the American Academy of Arts and Sciences*, vol. 80.

1951d. "Two Dogmas of Empiricism." *Philosophical Review*, vol. 60. Reprinted in Quine 1953a, Benacerraf and Putnam 1964, and Feigl, Sellars, and Lehrer 1972.

1952. "On an Application of Tarski's Theory of Truth." *Proceedings of the National Academy of Sciences*, vol. 38. Reprinted in Quine 1966a.

1953a. *From a Logical Point of View: 9 Logico-Philosophical Essays*. Cambridge: Harvard University Press; 2nd ed., 1961.

1953b. "On Mental Entities." In *Contributions to the Analysis and Synthesis of Knowledge. Proceedings of the American Academy of Arts and Sciences*, vol. 80. Reprinted in Quine 1966b.

1955. "On Frege's Way Out." *Mind*, n.s. vol. 64. Reprinted in Quine 1966a.

1958. "Speaking of Objects." *Proceedings and Addresses of the American Philosophical Association*, vol. 31.

1960. *Word and Object*. Cambridge and New York: Technology Press of MIT and Wiley; London: Chapman.

1962. "Carnap and Logical Truth." In *Logic and Language: Studies Dedicated to Professor Rudolf Carnap on the Occasion of His Seventieth Birthday*. Edited by B. H. Kazemier and D. Vuysje. Dordrecht: Reidel. Reprinted in Schilpp 1963, Quine 1966b, Feigl, Sellars, and Lehrer 1972, and in this volume.

1964. "Truth by Convention." In Benacerraf and Putnam 1964. Reprint with some changes of Quine 1936. Reprinted in this volume.

1966a. *Selected Logic Papers*. New York: Random House.

1966b. *The Ways of Paradox and Other Essays*. New York: Random House; 2nd rev. ed., Cambridge, Mass., and London: Harvard University Press, 1976.

Ramsey, F. P. 1926a. "The Foundations of Mathematics." *Proceedings of the London Mathematical Society* (1925), 2nd ser. vol. 25, pt. 5. Reprinted in Ramsey 1931.

1926b. "Mathematical Logic." *Mathematical Gazette*, vol. 13. Reprinted in Ramsey 1931.

1926c. "Universals and the 'Method of Analysis.'" In *Methods of Analysis. Aristotelian Society Supplementary Volume 6*. Reprinted in Ramsey 1931.

1931. *The Foundations of Mathematics and Other Logical Essays*. Edited by R. B. Braithwaite. New York: Harcourt, Brace; London: Routledge and Kegan Paul.

Reichenbach, H. 1944. "Bertrand Russell's Logic." In *The Philosophy of Bertrand Russell*. Edited by P. A. Schilpp. Evanston and Chicago: Northwestern University Press; 3rd ed., New York: Tudor, 1951.

Reinhardt, W. N. 1974a. "Remarks on Reflection Principles, Large Cardinals, and Elementary Embeddings." In *Axiomatic Set Theory. Proceedings of Symposia in Pure Mathematics*, vol. 13, pt. 2. Providence, R.I.: American Mathematical Society.

1974b. "Set Existence Principles of Shoenfield, Ackermann, and Powell." *Fundamenta Mathematicae*, vol. 84.

Resnik, M. D. 1966. "On Skolem's Paradox." *Journal of Philosophy*, vol. 63.

Robinson, A. 1965. "Formalism 64 [1964]." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1964 International Congress*. Edited by Y. Bar-Hillel. Amsterdam: North-Holland.

1973. "Mathematical Problems." *Journal of Symbolic Logic*, vol. 38.

Rogers, H., Jr. 1968. *Theory of Recursive Functions and Effective Computability*. New York: McGraw-Hill.

Rosser, J. B. 1936. "Constructibility as a Criterion for Existence." *Journal of Symbolic Logic*, vol. 1.

1939. "An Informal Exposition of Proofs of Gödel's Theorems and Church's Theorem." *Journal of Symbolic Logic*, vol. 4. Reprinted in Davis 1965.

Russell, B. A. W. 1903. *The Principles of Mathematics*. Cambridge University Press; 2nd ed., London: Allen and Unwin, 1937.

1905. "On Denoting." *Mind*, n.s. vol. 14. Reprinted in Feigl and Sellars 1949 and Russell 1956.

1906a. "Les Paradoxes de la logique." *Revue de métaphysique et de morale*, vol. 14.

1906b. "The Theory of Implication." *American Journal of Mathematics*, vol. 28.

1907. "On Some Difficulties in the Theory of Transfinite Numbers and Order Types." *Proceedings of the London Mathematical Society*, 2nd ser. vol. 4.

1908. "Mathematical Logic as Based on the Theory of Types." *American Journal of Mathematics*, vol. 30. Reprinted in Russell 1956 and van Heijenoort 1967.

1914. *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*. Chicago: Open Court; 2nd ed., Norton, 1929.

1918-19. "The Philosophy of Logical Atomism," parts 1-7. *Monist*, vols. 28 and 29. Reprinted in Russell 1956.

1919. *Introduction to Mathematical Philosophy*. London: Allen and Unwin; New York: Macmillan; 2nd ed., Allen and Unwin, 1920. Chapters 1, 2, and 18 reprinted in Benacerraf and Putnam 1964 and in this volume.

1940. *An Inquiry into Meaning and Truth*. London: Allen and Unwin; New York: Norton; reprinted, Allen and Unwin, 1943 and 1951.

1956. *Logic and Knowledge: Essays, 1901-1950*. Edited by R. C. Marsh. London: Allen and Unwin; New York: Macmillan.

1959. *My Philosophical Development*. New York: Simon and Schuster.

Ryle, G. 1946. "Symposium: Why Are the Calculuses of Logic and Mathematics Applicable to Reality?" part 1. In *Logic and Reality. Aristotelian Society Supplementary Volume 20*.

1949. "Discussion: *Meaning and Necessity*," a review of Carnap 1947. *Philosophy: The Journal of the Royal Institute of Philosophy*, vol. 24.

Rynin, D. 1956. "The Dogma of Logical Pragmatism." *Mind*, n.s. vol. 65.

Saaty, T. L., ed. 1963-5. *Lectures on Modern Mathematics*, 3 vols. New York: Wiley.

Scheffler, I., and Chomsky, N. 1959. "What Is Said to Be." *Proceedings of the Aristotelian Society* (1958-9), n.s. vol. 59.

Schilpp, P. A., ed. 1941. *The Philosophy of Alfred North Whitehead*. Evanston and Chicago: Northwestern University Press; 2nd ed., New York: Tudor, 1951.

1944. Ed. *The Philosophy of Bertrand Russell*. Evanston and Chicago: Northwestern University Press; 3rd ed., New York: Tudor, 1951.

1963. Ed. *The Philosophy of Rudolf Carnap*. La Salle, Ill.: Open Court.

Schlick, M. 1918. *Allgemeine Erkenntnislehre*. Berlin: Springer; 2nd ed., 1925.

1932. "Positivismus und Realismus." *Erkenntnis*, vol. 3. Reprinted in Schlick 1938. Translated as "Positivism and Realism" by D. Rynin in Ayer 1959.

1938. *Gesammelte Aufsätze, 1926-36*. Vienna: Gerold.

Schröder, E. 1873. *Lehrbuch der Arithmetik und Algebra für Lehrer und Studirende*. Leipzig: Teubner.

Schröter, K. 1973. "Interpretation der intuitionistischen Logik mit Hilfe des Beweisbarkeitsbegriffs." *Ajatus Suomen Filosofisen Yhdistyksen Vuosikiria*, vol. 35.

Schütte, K. 1960a. *Beweistheorie. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen*, vol. 103. Berlin: Springer.

1960b. "Syntactical and Semantical Properties of Simple Type Theory." *Journal of Symbolic Logic*, vol. 25.

1968. *Vollständige Systeme modaler und intuitionistischer Logik*. Heidelberg: Springer.

Schwartz, S. P., ed. 1977. *Naming, Necessity, and Natural Kinds*. Ithaca, N.Y.: Cornell University Press.

Scott, D. 1961. "Measurable Cardinals and Constructible Sets." *Bulletin de L'Académie Polonaise des Sciences, Série des Sciences Mathématiques, Astronomiques et Physiques*, vol. 9.

1968. "Extending the Topological Interpretation to Intuitionistic Analysis." *Compositio Mathematica*, vol. 20.

1970. "Extending the Topological Interpretation to Intuitionistic Analysis II." In *Intuitionism and Proof Theory. Proceedings of the Summer Conference at Buffalo, N.Y., 1968*. Edited by A. Kino, J. Myhill, and R. Vesley. Amsterdam: North-Holland.

Searle, J. R. 1958. "Russell's Objections to Frege's Theory of Sense and Reference." *Analysis*, vol. 18.

Sellars, W. 1949. "Acquaintance and Description Again." *Journal of Philosophy*, vol. 46.

Sheffer, H. M. 1913. "A Set of Five Independent Postulates for Boolean Algebras, with Applications to Logical Constants." *Transactions of the American Mathematical Society*, vol. 14.

Shoenfield, J. R. 1967. *Mathematical Logic*. Reading, Mass.: Addison-Wesley.

Shwayder, D. S. 1961. *Modes of Referring and the Problem of Universals: An Essay in Metaphysics. University of California Publications in Philosophy*, vol. 35. Berkeley: University of California Press.

Sierpiński, W. 1934a. *Hypothèse du Continu. Monografie Matematyczne*, vol. 4. Warsaw: Z Subwencji Funduszu Kultury Narodowej: 2nd ed., New York: Chelsea, 1956.

1934b. "Sur une extension de la notion de l'homéomorphie." *Fundamenta Mathematicae*, vol. 22.

1935a. "Sur deux ensembles linéaires singuliers." *Annali della R. Scuola Normale Superiore di Pisa: Scienze Fisiche e Matematiche*, 2nd ser. vol. 4.

1935b. "Sur une hypothèse de M. Lusin." *Fundamenta Mathematicae*, vol. 25.

1951. "Sur quelques propositions concernant la puissance du continu." *Fundamenta Mathematicae*, vol. 38.

Sierpiński, W., and Tarski, A. 1930. "Sur une propriété caractéristique des nombres inaccessibles." *Fundamenta Mathematicae*, vol. 15.

Sikorski, R. 1951. "A Characterization of Alephs." *Fundamenta Mathematicae*, vol. 38.

Skolem, T. 1923. "Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre." *Matematikerkongressen i Helsingfors den 4–7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse.* Helsinki: Akademiska Bokhandeln. Translated as "Some Remarks on Axiomatized Set Theory" by S. Bauer-Mengelberg in van Heijenoort 1967.

1934. "Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen." *Fundamenta Mathematicae*, vol. 23.

1970. *Selected Works in Logic.* Edited by J. E. Fenstad. Oslo: Universitetsforlaget.

Skyrms, B. 1967. "The Explication of '*X* Knows that *p*'." *Journal of Philosophy*, vol. 64.

Smart, J. J. C. 1950. "Whitehead and Russell's Theory of Types." *Analysis*, vol. 10.

Smorynski, C. 1977. "The Incompleteness Theorems." In *Handbook of Mathematical Logic.* Edited by J. Barwise. *Studies in Logic and the Foundations of Mathematics*, vol. 90. Amsterdam and New York: North-Holland.

Spector, C. 1957. "Recursive Ordinals and Predicative Set Theory." In vol. 3, *Summaries of Talks Presented at the Summer Institute for Symbolic Logic (Cornell University, 1957)*, 3 vols., duplicated; 2nd ed., Princeton, N.J.: Communications Research Division, Institute for Defense Analyses, 1960.

1962. "Provably Recursive Functionals of Analysis: A Consistency Proof of Analysis by an Extension of Principles Formulated in Current Intuitionistic Mathematics." In *Recursive Function Theory. Proceedings of Symposia in Pure Mathematics*, vol. 5. Providence, R.I.: American Mathematical Society.

Steiner, M. 1973. "Platonism and the Causal Theory of Knowledge." *Journal of Philosophy*, vol. 70.

1975. *Mathematical Knowledge.* Ithaca, N.Y.: Cornell University Press.

Strawson, P. F. 1950. "On Referring." *Mind*, n.s. vol. 59. Reprinted in Feigl, Sellars, and Lehrer 1972.

1952. *Introduction to Logical Theory.* London: Methuen; New York: Wiley.

1956. "Singular Terms, Ontology, and Identity." *Mind*, n.s. vol. 65.

1959. *Individuals.* London: Methuen.

Study, E. 1914. *Die realistische Weltansicht und die Lehre vom Raume.* Brunswick: Vieweg.

Tait, W. W. 1965a. "Functionals Defined by Transfinite Recursion." *Journal of Symbolic Logic*, vol. 30.

1965b. "Infinitely Long Terms of Transfinite Type." In *Formal Systems and Recursive Functions: Proceedings of the Eighth Logic Colloquium* (Oxford, 1963). Edited by J. N. Crossley and M. A. E. Dummett. Amsterdam: North-Holland.

Takeuti, G. 1954. "Construction of the Set Theory from the Theory of Ordinal Numbers." *Journal of the Mathematical Society of Japan*, vol. 6.

Takeuti, G., and Zaring, W. M. 1971. *An Introduction to Axiomatic Set Theory.* New York: Springer.

Tarski, A. 1925. "Quelques théorèmes sur les alephs." *Fundamenta Mathematicae*, vol. 7.

1930–1. "O pojęciu prawdy w odniesieniu do sformalizowanych nauk dedukcyjnych." *Ruch Filozoficzny*, vol. 12. Translated as Tarski 1935b.

1934. "Z badań metodologicznych nad definiowalnośćą terminów." *Przegląd Filozoficzny*, vol. 37. Translated as Tarski 1935a.

1935a. "Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe." *Erkenntnis*, vol. 5. Translated as "Some Methodological Investigations on the Definability of Concepts" in Tarski 1956.

1935b. *Der Wahrheitsbegriff in den formalisierten Sprachen. Studia Philosophica*, vol. 1. Translated as "The Concept of Truth in Formalized Languages" in Tarski 1956.

1936. "O pojciu wynikania logicznego." *Przegląd Filozoficzny*, vol. 39. Translated as "On the Concept of Logical Consequence" in Tarski 1956.

1938. "Über unerreichbare Kardinalzahlen." *Fundamenta Mathematicae*, vol. 30.

1939. "On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth." *Journal of Symbolic Logic*, vol. 4.

1941. *Introduction to Logic and the Methodology of the Deductive Sciences.* New York: Oxford University Press.

1944. "The Semantic Conception of Truth and the Foundations of Semantics." *Philosophy and Phenomenological Research*, vol. 4. Reprinted in Feigl and Sellars 1949 and Linsky 1952.

1956. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938.* Translated by J. H. Woodger. Oxford: Clarendon Press. Includes Tarski 1930–1, 1934, and 1936.

1962. "Some Problems and Results Relevant to the Foundations of Set Theory." In *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress* (Stanford, Calif.). Edited by E. Nagel, P. Suppes, and A. Tarski. Stanford, Calif.: Stanford University Press.

Tarski, A., Mostowski, A., and Robinson, R. M. 1953. *Undecidable Theories.* Amsterdam: North-Holland.

Tharp, L. 1975. "Three Theorems of Metaphysics." Unpublished.

Thomson, J. F. 1960. "What Achilles Should Have Said to the Tortoise." *Ratio*, vol. 3.

1963. "On Some Paradoxes." In *Analytical Philosophy.* Edited by R. J. Butler. Oxford: Blackwell.

Troelstra, A. S. 1969. *Principles of Intuitionism.* Berlin, Heidelberg, and New York: Springer.
1973. Ed. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis. Lecture Notes in Mathematics,* no. 344. Edited by A. Dold and B. Eckmann. Berlin, Heidelberg, and New York: Springer.
1977a. *Choice Sequences: A Chapter of Intuitionistic Mathematics.* Oxford: Clarendon Press.
1977b. "Completeness and Validity for Intuitionistic Predicate Logic." In *Colloque International de Logique* (Clermont-Ferrand, 1975). *Colloques Internationaux du Centre National de la Recherche Scientifique,* no. 249. Paris: Éditions du Centre National de la Recherche Scientifique.
Turing, A. M. 1937. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society,* 2nd ser. vol. 42. Reprinted in Davis 1965.
1954. "Solvable and Unsolvable Problems." *Science News,* vol. 31.
Vacca, G. 1903. "La logica di Leibniz." *Rivista di matematica,* vol. 8.
van Heijenoort, J., ed. 1967. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931.* Cambridge: Harvard University Press.
Ville, F. 1971. "Decidabilité des formules existentielles en théorie des ensembles." *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences,* series A (Paris), vol. 272.
von Neumann, J. 1923. "Zur Einführung der transfiniten Zahlen." *Acta Litterarum ac Scientiarum: Sectio Scientiarum Mathematicarum,* vol. 1. Translated as "On the Introduction of Transfinite Numbers" by J. van Heijenoort in van Heijenoort 1967.
1925. "Eine Axiomatisierung der Mengenlehre." *Journal für die reine und angewandte Mathematik,* vol. 154. Translated as "An Axiomatization of Set Theory" by S. Bauer-Mengelberg and D. Føllesdal in van Heijenoort 1967.
1927. "Zur Hilbertschen Beweistheorie." *Mathematische Zeitschrift,* vol. 26.
1928. "Die Axiomatisierung der Mengenlehre." *Mathematische Zeitschrift,* vol. 27.
1929. "Über eine Widerspruchsfreiheitsfrage in der axiomatischen Mengenlehre." *Journal für reine und angewandte Mathematik,* vol. 160.
1931. "Die formalistische Grundlegung der Mathematik." *Erkenntnis,* vol. 2. Translated as "The Formalist Foundations of Mathematics" by E. Putnam and G. J. Massey in Benacerraf and Putnam 1964 and reprinted in this volume.
1961–3. *Collected Works,* 6 vols. Edited by A. H. Taub. Oxford and New York: Pergamon Press.
Waismann, F. 1928. "Die Natur des Reduzibilitätsaxioms." *Monatshefte für Mathematik und Physik,* vol. 35.
1938. "Ist die Logik eine deduktive Theorie?" *Erkenntnis,* vol. 7.
1946. "Are There Alternative Logics?" *Proceedings of the Aristotelian Society* (1945–6), n.s. vol. 46.
1949–53. "Analytic-Synthetic," parts 1–6. *Analysis,* vols. 10, 11, and 13.
1951. *Introduction to Mathematical Thinking.* Translated by J. Benac. New York: Ungar.

Wang, H. 1953. "What Is an Individual?" *Philosophical Review,* vol. 62.
1954. "Formalization of Mathematics." *Journal of Symbolic Logic,* vol. 19.
1955. "On Formalization." *Mind,* n.s. vol. 64.
1958. "Eighty Years of Foundational Studies." *Dialectica,* vol. 12.
1961. "Process and Existence in Mathematics." In *Essays on the Foundations of Mathematics, Dedicated to A. A. Fraenkel on His Seventieth Anniversary.* Edited by Y. Bar-Hillel, E. I. J. Poznanski, M. O. Rabin, and A. Robinson. Jerusalem: Magnes Press, Hebrew University.
1962. *A Survey of Mathematical Logic.* Peking: Science Press; Amsterdam: North-Holland, 1963.
1970. "Logic, Computation, and Philosophy." *L'Âge de la Science,* vol. 3.
1974. *From Mathematics to Philosophy.* London: Routledge and Kegan Paul. Chapter 6, "The Concept of Set," is reprinted in this volume.
1977. "Large Sets." In *Logic, Foundations of Mathematics, and Computability Theory. Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science* (London, Ontario, 1975), pt. 1. Edited by R. E. Butts and J. Hintikka. Dordrecht: Reidel.
Warnock, G. J. 1956. "Metaphysics in Logic." In *Essays in Conceptual Analysis.* Edited by A. Flew. London: Macmillan.
Weyl, H. 1918. *Das Kontinuum: Kritische Untersuchungen über die Grundlagen der Analysis.* Leipzig: Veit; reprinted, Berlin and Leipzig: Gruyter, 1932.
1921. "Über die neue Grundlagenkrise der Mathematik." *Mathematische Zeitschrift,* vol. 10.
1927. *Philosophie der Mathematik und Naturwissenschaft,* parts 1 and 2. *Handbuch der Philosophie,* nos. 4 and 5. Munich and Berlin: Oldenbourg. Translated as Weyl 1949.
1940. "The Mathematical Way of Thinking." *Science,* vol. 92.
1944. "David Hilbert and His Mathematical Work." *Bulletin of the American Mathematical Society,* vol. 50.
1946. "Mathematics and Logic: a brief survey serving as preface to a review of *The Philosophy of Bertrand Russell*" (Schilpp 1944). *American Mathematical Monthly,* vol. 53.
1949. *Philosophy of Mathematics and Natural Science.* Weyl 1927, revised and enlarged, based on a translation by O. Helmer. Princeton: Princeton University Press; reprinted, New York: Atheneum, 1963.
White, M. G. 1950. "The Analytic and the Synthetic: An Untenable Dualism." In *John Dewey, Philosopher of Science and Freedom: A Symposium.* Edited by S. Hook. New York: Dial Press. Reprinted in Linsky 1952.
White, N. P. 1974. "What Numbers Are." *Synthèse,* vol. 27.
Whitehead, A. N., and Russell, B. A. W. 1910–13. *Principia Mathematica,* 3 vols. Cambridge University Press; 2nd ed., 1925–7.
Wiggins, D. 1977. "Truth, Invention and the Meaning of Life." In *Proceedings of the British Academy* (London, 1976), vol. 62. London: Oxford University Press.
Wilder, R. L. 1952. *Introduction to the Foundations of Mathematics.* New York: Wiley.

Wittgenstein, L. 1921. *Logisch-philosophische Abhandlung. Annalen der Naturphilosophie*, vol. 14. Reprinted and translated as Wittgenstein 1922 and 1961c.

1922. *Tractatus Logico-Philosophicus.* A translation of Wittgenstein 1921 with German and English on opposite pages and an introduction by B. Russell. London: Routledge and Kegan Paul. Reprinted with a new translation as Wittgenstein 1961c.

1929. "Some Remarks on Logical Form." In *Knowledge, Experience and Realism. Aristotelian Society Supplementary Volume 9.*

1953. *Philosophical Investigations.* Translated by G. E. M. Anscombe. Oxford: Blackwell.

1956. *Remarks on the Foundations of Mathematics.* Translated by G. E. M. Anscombe. Edited by G. H. von Wright, R. Rhees, and G. E. M. Anscombe. Oxford: Blackwell; New York: Macmillan. Selections from *Remarks* were reprinted in Benacerraf and Putnam 1964.

1961a. *The Blue and Brown Books.* Oxford: Blackwell.

1961b. *Notebooks 1914–1916.* Translated by G. E. M. Anscombe. Edited by G. H. von Wright and G. E. M. Anscombe. Oxford: Blackwell.

1961c. *Tractatus Logico-Philosophicus.* A translation of Wittgenstein 1921 by D. F. Pears and B. F. McGuinness, with an introduction by B. Russell. London: Routledge and Kegan Paul; New York: Humanities Press.

Wood, O. P. 1961. "Symposium: On Being Forced to a Conclusion II." *Aristotelian Society Supplementary Volume 35.*

Young, J. W. 1911. *Lectures on the Fundamental Concepts of Algebra and Geometry.* New York: Macmillan.

Zermelo, E. 1904. "Beweis, dass jede Menge wohlgeordnet werden kann." *Mathematische Annalen*, vol. 59. Translated as "Proof That Every Set Can Be Well-Ordered" by S. Bauer-Mengelberg in van Heijenoort 1967.

1908. "Untersuchungen über die Grundlagen der Mengenlehre I." *Mathematische Annalen*, vol. 65. Translated as "Investigations in the Foundations of Set Theory I" by S. Bauer-Mengelberg in van Heijenoort 1967.

1909. "Sur les ensembles finis et le principe de l'induction complète." *Acta Mathematica*, vol. 32.

1929. "Über eine Widerspruchsfreiheitsfrage in der axiomatischen Mengenlehre." *Journal für reine und angewandte Mathematik*, vol. 160.

1930. "Über Grenzzahlen und Mengenbereiche." *Fundamenta Mathematicae*, vol. 16.