THE JOURNAL OF SYMBOLIC LOGIC
Volume 19, Number 4, Dec. 1954

THE FORMALIZATION OF MATHEMATICS[1]

HAO WANG

1. Original sin of the formal logician. Zest for both system and objectivity is the formal logician's original sin. He pays for it by constant frustrations and by living ofttimes the life of an intellectual outcaste. The task of squeezing a large body of stubborn facts into a more or less rigid system can be a painful one, especially since the facts of mathematics are among the most stubborn of all facts. Moreover, the more general and abstract we get, the farther removed we are from the raw mathematical experience. As intuition ceases to operate effectively, we fall into many unexpected traps. The formal logician gets little sympathy for his frustrations. He is regarded as too rigid by his philosophical colleagues and too speculative by his mathematical friends. The life of an intellectual outcaste may be a result partly of temperament and partly of the youthfulness of the logic profession. The unfortunate lack of wide appeal of logic may, however, be prolonged partly on account of the fact that very little of the well-established techniques of mathematics seems applicable to the treatment of serious problems of logic.

The axiomatic method is well suited to provide results which are both exact and systematic. How attractive would it be if we could get an axiom system in which all the axioms and deductions were intuitively clear and all theorems of mathematics were provable? Such a system would undoubtedly satisfy Descartes who admits solely intuition and deduction, which are, for him, the only "mental operations by which we are able, wholly without fear of illusion, to arrive at the knowledge of things." Indeed, according to Descartes, intuition and deduction "are the most certain routes to knowledge, and the mind should admit no others. All the rest should be rejected as suspect of error and dangerous."[2]

2. Historical perspective. Euclid's unification of masses of isolated discoveries in Greek elementary geometry was undoubtedly the first impressive success in systematizing mathematics. In a way it came about quite naturally. Of the four hundred sixty-five propositions in Euclid's ***Elements***, some are quite obvious to our geometrical intuition, some are not

Received June 6, 1954. An address delivered, by invitation of the Program Committee, to a meeting of the Association for Symbolic Logic at Rochester, New York, on December 29, 1953.

[1] I wish to thank the referee and Professor Max Black for useful comments on an earlier draft of this paper.

[2] ***The rules for the direction of the mind,*** Rule III.

so obvious. Confronted with the existing proofs of the less obvious by the more obvious, it was natural to ask how the various theorems are interrelated. Then it became mainly a matter of perseverance and acuteness to get the theorems arranged the way Euclid actually did arrange them.

There are, however, several points worth remarking. In the first place, systematization calls for more than the ability of a good librarian. For example, it was not until the nineteenth century that Pasch first formulated axioms concerning the concept "between" which had been tacitly assumed but not explicitly stated in Euclid. Moreover, a field has often to be developed very thoroughly before it is ripe for a systematic and rigorous organization. The history of the calculus illustrates this point clearly: founded in the seventeenth century, rapidly expanded in the eighteenth, the calculus got acceptable foundations only in the nineteenth century and even today logicians generally have misgivings on the matter or, like Weyl, still think that analysis is built on sand.

During the nineteenth century, the attempts to found analysis on a reliable basis went generally under the caption "arithmetization of analysis." It is well known that Cauchy, Weierstrass, Dedekind, Cantor all made important contributions to this program. Indeed, their results were so well received among mathematicians that in 1900, Poincaré asserted: "Today there remain in analysis only integers and finite or infinite systems of integers, interrelated by a net of relations of equality or inequality. Mathematics, as we say, has been arithmetized. ...We may say today that absolute rigour has been attained."[3]

If by "arithmetization" is meant merely the elimination of geometrical intuition, then the success is hardly disputable. If, on the other hand, by "arithmetization" is meant a reduction of analysis to a theory of integers, then the matter becomes more involved because not only integers but also "finite or infinite systems of integers" are needed. Nowadays it would be more customary to refer to these "systems" as sets or classes. What is accomplished is not the founding of analysis on the theory of integers alone, but rather on the theory of integers plus the theory of sets. Therefore, the problem of getting a satisfactory theory of real numbers and real functions is not solved but shifted in a large part to the problem of finding a satisfactory theory of sets. And, as we know, to get a wholesome set theory is no small matter.

3. What is a set? More explicitly, to get a rigorous basis of the calculus, an exact theory of the continuum is needed. Since real numbers can be regarded as arbitrary sets of rational numbers or positive integers which

---

[3] *Du rôle de l'intuition et de la logique en mathématiques*, **Compte-rendu du IIième Congrès International des Mathématiciens**, 1900, Paris (1902), pp. 200–202.

satisfy a few very broad conditions, this means that an exact development of the calculus logically calls for a general theory of sets.

There is also another way in which problems of ordinary mathematics should have led in the nineteenth century to queries as to what a set is. There were frequent occasions to consider arbitrary curves or arbitrary functions of real numbers. For example, is an arbitrary function representable by a trigonometric series? What functions are integrable? Many serious mathematicians were busy with such problems. Yet to answer these questions, it would appear prerequisite to have a pretty good idea of what an arbitrary function or set is.

It is of interest to note that as an historical fact mathematicians often speak of arbitrary functions and arbitrary curves when they have no precise definition of these notions and actually have in mind only certain special functions and special curves. The great discrepancy between the really arbitrary and the moderate arbitrariness which is actually needed in living mathematics explains the possibility of various basically different systems of set theory which compete to provide the true foundations of mathematics.

In the nineteenth century nobody paused to supply an exact definition of the notion of arbitrary set or arbitrary function. To do so would have required a thorough examination of all the means of definition at their disposal. There was at that time neither reason to suppose this work necessary nor enough advance preparation for carrying it out. Only after the discovery of paradoxes around 1900 was it realized that not all apparent laws or definitions could define sets and that some restriction on the permissible means of defining sets was necessary.

The historical course of events was different from the logical process of descending from the more abstract to the less general. Cantor did not have a general set theory to begin his investigations but was rather led to the study of point sets (sets of real numbers) by a comparatively more concrete problem which arose quite naturally from ordinary mathematics.

The problem is that of representing functions by trigonometric series which interested many a mathematician when Cantor began his research career around 1870. In trying to extend the uniqueness of representation to certain functions with infinitely many singular points, he was led to the notion of a derived set which not only marked the beginning of his study of the theory of point sets but led him later on to the construction of transfinite ordinal numbers.

Such historical facts ought to help combat the erroneous impression that Cantor invented, by one stroke of genius, a whole theory of sets which was entirely isolated from the main stream of mathematics at his time. In addition, it may be interesting to recall that mathematicians such as Heine and Dedekind were familiar with the problems which Cantor treated in his set theory and quite capable of handling them had they wished to. Indeed, du Bois-Reymond discovered independently of Cantor the notion of derived set, as well as the notion of derived set of infinite orders which led Cantor to the transfinite ordinals of the second number class. Moreover, du Bois-Reymond anticipated Cantor[4] by about twenty years in using the diagonal argument now generally attributed to Cantor. The main reason why Cantor has been so much more influential is probably his ability to free himself gradually from applications and develop the theory of sets more and more for its own sake. Only in thus generalizing and following up logical conclusions everywhere, did Cantor become the founder of set theory.

4. The indenumerable and the impredicative. The notions of denumerability and well-ordering were of central importance for Cantor: the former is the pillar of his theory of cardinal numbers, the latter of his theory of ordinal numbers.

In inventing set theory, the two most remarkable jumps which Cantor made were: the invention of transfinite ordinal numbers of his second number class, and the use of indenumerable and impredicative sets. The first is now known to be harmless and useful (especially in certain metamathematical considerations), while the second remains a mystery which has shed little light on any problems of ordinary mathematics. There is no clear reason why mathematics could not dispense with impredicative or absolutely indenumerable sets.

Cantor gives two proofs of the indenumerability of real numbers and one proof of the indenumerability of his second number class. All these proofs make use of impredicative sets.

Since not everybody is familiar with the nature of impredicative definition, it may be worthwhile to pause and review the well-known diagonal argument for proving the indenumerability of real numbers.

To prove this, it is, as we know, sufficient to prove that the set M of all sets of positive integers is not denumerable. Cantor's proof for this is as follows. Suppose there were a one-to-one correspondence $f(x, k)$ or $x = f(k)$ between M and the set P of positive integers, so that for every given positive integer k_0, there is a set $f(k_0)$ in M which is the image of k_0. For each positive integer k, either k belongs to its image $f(k)$ or not. Consider the set N of all positive integers k such that k does not belong to $f(k)$. N would be a set of positive integers and therefore a member of M. By hypothesis, there would be a positive integer n such that N is $f(n)$. Either n belongs to $f(n)$ or not. If n belongs to $f(n)$, then, by the definition of N, n does not belong to $f(n)$,

[4] See, e.g., P. du Bois-Reymond, *Über asymptotische Werte, infinitäre Approximationen und infinitäre Auflösung von Gleichungen*, **Mathematische Annalen**, vol. 8 (1875), pp. 363–414.

$f(n)$ being N. If N does not belong to $f(n)$, then by the definition of N again, n belongs to $f(n)$. Hence, we obtain a contradiction. It follows that given any one-to-one correspondence between positive integers and sets of positive integers, we can always find a set of positive integers which is different from all the sets already enumerated.

We can make a number of different comments on the arguments. From the proof, it certainly follows that given any law which enumerates sets of positive integers

$$x_1, x_2, x_3, \ldots,$$

we can find a set x which is different from every one of the above. Moreover, given x and the sets $x_1, x_2, \ldots$, we can also find a law which enumerates $x_1, x_2, \ldots$ together with x:

$$x, x_1, x_2, x_3, \ldots;$$

but then there is another set y which is different from all these. Then we can also find another sequence which includes y and all terms of the previous sequence. And so on.

From the fact that no enumeration can exhaust all sets of positive integers, Cantor infers that the set of all sets of positive integers is absolutely indenumerable. In order to justify this inference, we have to assume that there is a set which includes all sets of positive integers, or that there is a law which defines a set that includes all sets of positive integers. Constructive set theory refuses to recognize any such set or any such law. While the constructive viewpoint accepts totalities of laws relative to different stages of construction, it rejects a closed totality which excludes possibilities of further construction.

If, however, we do allow that there is a set M of *all* sets of positive integers, then the above argument shows that such a set M is absolutely indenumerable. Moreover, we then see that the argument uses what is known as an impredicative definition to define the set N of positive integers. Thus, by definition, k belongs to N if and only if *there exists* a set K in M such that K is $f(k)$ and k does not belong to K. This is impredicative because in defining N, we make use of the totality M which contains N as a member. Thus, in order to determine whether a given k_0 belongs to N, we have, so to speak, to check through all members of M (including N itself) to see whether some one of them is $f(k_0)$ and yet does not contain k_0 as member. Hence, in order to define N, N must already be there. This is clearly inacceptable from a constructive viewpoint according to which the set only comes into being by a definition. Only if we take the sets as somehow existing before we say anything about them, can we accept such definitions: and then, not as definitions, but as descriptions of properties which sets possess by themselves, or as directions for picking suitable sets from a huge ocean containing all sorts of familiar, as well as curious, fish.

The situation is quite similar with Cantor's proof that the second number class is not denumerable.

It is no accident that all proofs of absolute indenumerability use impredicative sets. Indeed, it is at least intuitively plausible that all predicative sets are denumerable. We may also recall that Russell's paradox was first obtained from analysis of Cantor's diagonal argument and uses an impredicative set, too.

5. The limitations upon formalization. A satisfactory axiomatization of set theory seems to be the most hopeful way of carrying out the ambitious program of systematizing all mathematics. At the beginning of the century the discovery of paradoxes plus the popularity of axiomatic method in geometry and arithmetic led quite naturally to attempts to construct axiom systems for set theory in which as much of Cantor's "naive" theory as possible, but of course none of the paradoxes, is to be derived. As a result, we possess today a number of axiom systems for set theory.

In order to formalize mathematics in axiomatic set theory it is customary, on account of the great diversity of content, to make certain preliminary representations or reductions. For example, geometrical points can be represented by real numbers, functions and relations can be construed as sets (of couples, etc.), numbers can be identified with certain special sets, and so on. With these reductions, it is often asserted and believed that each of several standard axiom systems of set theory is adequate to the development of all mathematics. For example, the system of ***Principia mathematica,*** or the system constructed by Zermelo and extended by his successors.

The actual derivation of mathematics from any such system is long and tedious. It is practically impossible to verify conclusively any claim that this and that branch of mathematics, with all their details, are derivable in such a system. It is also hard to refute such a claim for that would require the discovery of some premise or principle of inference, which has so far been tacitly assumed but unrecognized.

There are, nonetheless, at least three rather general objections to claims of this sort. In the first place, none of these standard systems is known to be free from contradictions. In the second place, each system can easily be expanded, for instance, by adding a higher level of sets, to a new system which at the same time contains new theorems and yet can be proved to be no less reliable than the original. Indeed, a proof can be given in each case for the conclusion that the extended system is consistent, provided the original is. Moreover, the extended system can again be similarly extended; the process of such extensions can be continued indefinitely. A closely related difficulty is that all these systems are subject to the Gödel imcompleteness. A third objection is that none of these systems can supply all the set theory as originally constructed by Cantor. Specifically, this has to do with Cantor's notion of indenumerable sets and his use of impredicative definitions. While Cantor asserts the existence of sets which are absolutely indenumerable, any of these axiom systems can supply only sets which are indenumerable relative to the means of expression in the system. Such sets are denumerable in the absolute sense, as it is possible to enumerate the elements of each by talking *about* the original axiom system.

There are, largely as a result of the first objection, numerous attempts to construct artificial systems which are both demonstrably consistent and also adequate to the development of the "principal" (or "useful") parts of mathematics. Most of these systems modify our basic logical principles such as the law of excluded middle and the principle of extensionality (for sets), and it is not easy to become familiar with them. So far as I know, none of these has been accepted widely.

The attitude toward the second and the third objections is usually either one of indifference or one of resignation. These objections, it is argued, need not be taken seriously, since what we have is already sufficient for all ordinary purposes and the creation of new inadequacies by considering the system as given is quite idle. Others contend that these difficulties are the price which we have to pay for using a formal language or using an axiom system.

In what follows an approach will be suggested in outline which is both natural and not subject to the above three objections.

6. A constructive theory. This is not the place to describe the formal details of the constructive theory which is claimed to possess all the wonderful properties of naturalness, adequacy, and demonstrable consistency. The theory will only be roughly sketched and an attempt will be made to make it appear plausible that the theory can do the things which it is supposed to do. I possess a more exact treatment of the matter which I hope will in the near future become available for scrutiny to those who are interested.

The system or theory will be denoted by the capital Greek letter sigma Σ. It has in the lowest order (the 0-th order) a denumerable totality consisting of (say) all the positive integers or all the finite sets built up out of the empty set. In the first order are these same sets plus sets of them which can be defined by properties referring at most only to the totality of all sets of the 0-th order (or, in other words, by formulas which contain no bound variables of the first or a higher order). Similarly, for every positive integer n, the sets of order $n + 1$ include all sets of order n together with sets of them defined by properties referring at most only to the totality of all sets of the n-th order. The sets of order ω include all and only sets of the finite orders. For any ordinal number $\alpha + 1$, the sets of order $\alpha + 1$ are related to those of order α in the same way as the sets of order $n + 1$ to those of order n (n a nonnegative integer). For any ordinal number β which is the limit number of a monotone increasing sequence $\alpha_1, \alpha_2, \ldots$ of ordinals, the sets of order β are related to the sets of orders $\alpha_1, \alpha_2, \ldots$ in the same way as the sets of order ω are related to those of finite orders. In short, sets of orders higher than 0 are constructed according to the Poincaré-Russell vicious-circle principle.

All the ordinal numbers which are used belong, of course, to what is known as Cantor's second number class. Moreover, we use only "constructive" ordinals. What a constructive ordinal number is presents an interesting and difficult problem. A simple and straightforward characterization of the totality of constructive ordinals is likely to get into the difficulty that the diagonal argument would produce a new ordinal number which should again be regarded as constructive. Let us assume for the moment that a suitable notion of constructive ordinals is given. Longer discussion of the topic will be included in a later section. We observe merely that in any case the usual ordinal numbers of Cantor's second number class, such as ω^2, the ε-numbers, all are constructive ordinals.

The axioms of the theory Σ can be briefly described as follows. All axioms and rules of inference of the standard quantification theory (predicate calculus) hold with regard to sets of each order. Set terms or abstracts of different orders are included in the primitive notation. Two sets x_α and y_β are equal or $x_\alpha = y_\beta$ if and only if they have the same extension; or, more exactly, if $\alpha \geq \beta$, every set z_α belongs either to both x_α and y_β or to none. Special axioms of the theory are:

A. Identity: for every γ, $\gamma \geq \alpha$, $\gamma \geq \beta$, if $x_\alpha = y_\beta$ and $x_\alpha \in z_\gamma$, then $y_\beta \in z_\gamma$.

B. Infinite summation: for every limiting ordinal number α, if $\beta < \alpha$, then for every x_β, there is y_α, such that $y_\alpha = x_\beta$.

C. Abstraction: for every formula $F(x_\beta)$, every y_β belongs to $\hat{x}_\beta F(x_\beta)$ if and only if $F(y_\beta)$.

D. Foundation: if x_α is not empty, then there is some y_α such that $y_\alpha \in x_\alpha$, and y_α and x_α have no common member.

E. Bounded order: if $x \in y$ and y is not of higher order than x, then there exists a set z of lower order than y such that $x = z$.

F. Limitation: see III in the next section.

The ideas employed in the construction of the theory Σ are not new. Russell, Weyl, Chwistek, Lorenzen all have developed mathematics along somewhat similar lines. I shall not enter into detailed comparison of the present approach with works of these authors, except merely to remark that

Russell and Weyl do not even claim adequacy of their systems for the development of analysis, that I have not been able to understand Chwistek, and that Lorenzen has little regard for formalization. Many of the conclusions obtained by using the theory Σ are very much the same as what Lorenzen arrives at from a somewhat informal approach.

The theory Σ is not exactly a logistic system but rather a system schema. It is the union of all formal systems Σ_α, where α is an arbitrary constructive ordinal, and Σ_α deals with all and only those sets which are of order α or less. By referring to these partial systems Σ_α, we shall be able to make many quite exact statements about the comprehensive theory Σ.

7. The denumerability of all sets. One peculiarity about the theory Σ is that all sets of Σ are enumerable in Σ. Indeed, it is possible to enumerate all sets of any given order α by a function of order $\alpha + 2$. By using standard methods for giving Tarski truth definitions and constructions employed by Bernays in his demonstration of a general class metatheorem from a finite number of axioms, we can prove, at least for α not too large (say less than ω^2), the following two results:

I. For each α, we can find a function E_α of order $\alpha + 2$, such that E_α enumerates all sets of order α; or, in other words, the domain of E_α is the set of all positive integers and its range is the universal set V_α consisting of all sets of order α.

II. For each α, we can find a truth definition of Σ_α in $\Sigma_{\alpha+2}$ and formalize a consistency proof of Σ_α in $\Sigma_{\alpha+2}$ (i.e., prove $\mathrm{Con}(\Sigma_\alpha)$ in $\Sigma_{\alpha+2}$).

Incidentally, it may be of interest to compare these with similar results on ordinary predicate calculi in which impredicative definitions are allowed. For example, consistency of the predicate calculus of type n can be proved in that of type $n + 1$, by the use of impredicative sets; while here we have to prove the consistency of Σ_α in $\Sigma_{\alpha+2}$. Both in proving the consistency of Σ_α and in enumerating V_α, we have to use sets which take sets of order α as members but are defined with the help of bound variables of order $\alpha + 1$. While these sets are impredicative sets of type $\alpha + 1$ according to their members, they are sets of order $\alpha + 2$ according to their definitions. That is why in both I and II we have to use the order $\alpha + 2$ instead of $\alpha + 1$.

Using the functions E_α, we are able to state powerful axioms of limitation which stipulate that in the theory Σ we recognize no sets other than those explicitly enumerated by the functions E_α (compare also Chwistek's axiom of enumerability and Fitch's hypothesis of similarity[5]):

---

[5] Frederic B. Fitch, *The hypothesis that infinite classes are similar*, this Journal, vol. 4 (1939), pp. 159–162.

Leon Chwistek, *Über die Hypotheses der Mengenlehre*, ***Mathematische Zeitschrift,*** vol. 25 (1926), pp. 439–473.

III. For each order α and each set x_α, there is a positive integer m such that $E_\alpha(m)$ is x_α.

From these axioms of limitation, it becomes possible to prove general theorems on all sets of each order by using mathematical induction. Since E_α well-orders all sets of order α, certain axioms of choice can be proved. Moreover, it is clearly also possible to enumerate all sets of order α which are ordinal numbers. It follows that we can, in $\Sigma_{\alpha+2}$, find a one-to-one correspondence between all sets of order α and all sets of order α which are ordinal numbers. One can also prove, by the diagonal argument, that no such correlation exists in Σ_α itself. Therefore, the continuum hypothesis (viz. the hypothesis that the set of all sets of order α has a different cardinality than the set of all sets of order α which are ordinal numbers) is provable or refutable according as whether equi-cardinality is defined by the existence of a correlation of order $\alpha + 2$ or one of order α. In short, we have:

IV. Axioms of choice are provable in Σ.

V. Certain forms of the continuum hypothesis are provable in Σ; certain other forms are refutable in Σ.

Moreover, since $\mathrm{Con}(\Sigma_\alpha)$ is provable in $\Sigma_{\alpha+2}$, the Gödel undecidable propositions of each Σ_α are provable in $\Sigma_{\alpha+2}$. Hence, the only possible way to construct a Gödel proposition which is undecidable in Σ would be, so far as I can see, to find a sequence of increasing constructive ordinals $\alpha_1, \alpha_2, \ldots$ such that its limit is no longer a constructive ordinal and consider the union of $\Sigma_{\alpha_1}, \Sigma_{\alpha_2}, \ldots$; yet there is, so far as I can see, no apparent way to show that such a union is again a formal system or a system to which Gödel's constructions are applicable. We have, therefore:

VI. Gödel's famous constructions do not yield directly any propositions which are undecidable in the theory Σ.

It would be of interest to investigate whether there might not be some indirect means of constructing undecidable propositions in Σ.

Developing real numbers in some standard fashion[6], we can also prove standard theorems of classical analysis in the theory Σ. For example, we can prove, with pretty much the traditional arguments, the theorem of least upper bound, the Bolzano-Weierstrass theorem, and the Heine-Borel theorem.

This is actually not surprising since we do not hesitate, when necessary, to use sets of higher orders. For example, in the general case, the least upper bound of a set whose members are of order α is a set of order $\alpha + 1$. Both Weyl and Russell were aware of such possibilities but they found the use of higher orders objectionable. This and related points will, in the next section, be discussed at great length.

[6] E.g., as in Bernays' series of articles on axiomatic set theory in this JOURNAL, vol. 2 (1937); vol. 6 (1941); vol. 7 (1942); vol. 8 (1943); vol. 13 (1948); vol. 19 (1954).

Let me insert a few words on the use of indenumerable sets in measure theory. It is widely known that in measure theory there is a theorem stating that every denumerable set is of measure zero. It would seem that in the theory Σ where all sets are denumerable, the whole measure theory would collapse. In actuality, however, this is not the case because, although there are no sets which are absolutely indenumerable, for each a, there are sets of order a which are indenumerable by any functions of order a. And the notion of relative indenumerability is sufficient to provide us with sets of nonzero measures for measures defined on each given level.

Thus, let us recall the standard proof of the theorem. Given a denumerable set M of points

$$x_1, x_2, x_3, \ldots,$$

we can choose an arbitrarily small ϵ and cover each x_i by the interval from $x_i - \frac{\epsilon}{2^i}$ to $x_i + \frac{\epsilon}{2^i}$, then it is easy to see that the sum of these intervals is no greater than 2ϵ and that the measure of the original set is smaller than 2ϵ. Hence the set M has measure zero.

For each a, this proof can be carried out in Σ_a only if the given denumerable set M of points can be enumerated within Σ_a. It is perfectly possible for the same set to have measure zero in Σ_{a+1} and a nonzero measure in Σ_a.

I think this situation is completely satisfactory. We even get an explanation of the relation between the continuous and the discrete (i.e., denumerable). A set of points is continuous only relative to our knowledge or our power to isolate the numerous points. What is seen as continuous in a less powerful set theory becomes discrete as we come to use a richer set theory.

8. Consistency and adequacy. From the known consistency of the ramified theory of types it is natural to expect the consistency of the theory Σ. Indeed, the consistency of each system Σ_a can be proved similarly to that of the ramified type theory. Since the theory Σ is the union of all systems Σ_a, the consistency of Σ follows immediately.

We can prove the consistency of each system Σ_a by describing more in detail its intuitive model. The arguments are similar to Fitch's proof[7] of the consistency of the ramified theory of types.

Or, we can also give a proof-theoretic consistency proof of each system Σ_a. Such a proof is analogous to Lorenzen's and Schütte's finitist proofs[8] of the consistency of the ramified type theory.

[7] *The consistency of the ramified* Principia, this JOURNAL, vol. 3 (1938), pp. 140–149.

[8] PAUL LORENZEN, *Algebraische und logistische Untersuchungen über freie Verbände*, this JOURNAL, vol. 16 (1951), pp. 81–106; KURT SCHÜTTE, *Beweistheoretische Untersuchung der verzweigten Analysis*, **Mathematische Annalen**, vol. 124 (1951–52), pp. 123–147.

It is very difficult to explain exactly the sense in which Lorenzen's proof is finitist. But the following vaguely specified difference between it and Fitch's proof is clearly relevant.

Given any proof in the formal system of the ramified theory of types, Lorenzen's procedure can change it effectively into a different proof in which every line is more complicated than each of its premises (indeed, longer except for occasional deletion of repetitious terms). In other words, we get cumulative proofs. It is true that these proofs can contain potentially infinitely many propositions, e.g., in order to prove $(m)m \geq 0$, we use $0 \geq 0$, $1 \geq 0$, $2 \geq 0$, etc. Nonetheless, each application of the rule of infinite induction is determined unambiguously by the corresponding step of the original proof. Consequently, the actual applications of the rule of infinite induction are much less unmanageable than what the abstract statement of the rule itself would lead us to think.

In Fitch's proof by an intuitive model, the situation is more complex. It is not easy to see how we can get effectively, from a proof in the formal system, a corresponding proof in the intuitive model. For example, in order to justify that for every proposition p, 'p or not-p' belongs to the set of true propositions, we have to apply the law of excluded middle to the infinite set of true propositions and argue that either p belongs to the set, whence 'p or not-p' also belongs to it, or else p does not belongs to the set, whence not-p belongs to it and hence also 'p or not-p.' It is not possible to avoid altogether reference to the alternative that a proposition does not belong to the set of true propositions (i.e., belongs to the set of propositions which are not true). In other words, we cannot carry out the consistency proof without constantly 'going out of' the set of true propositions.

Naturally the consistency proofs for Σ involve peculiar features on account of special characteristics of the theory. Yet it is certainly reasonable to expect that these proofs can be carried out. Details will not be given here.

At several places we proved the required theorems only by ascending to higher orders: the axioms of choice, the reinterpretations of the continuum hypothesis, the least upper bound theorem, the Bolzano-Weierstrass theorem. For many people this seems to cut through rather than meet the original difficulties.

For example, Weyl mentions in ***Das Kontinuum*** (p. 23) the possibility of defining the least upper bound of a given bounded set of real numbers by a real number of a higher order, but discarded it immediately as philosophically unsatisfactory.

For a ramified analysis, an obvious difficulty which has often been mentioned is: since the least upper bound of a given set of real numbers is a real number of a higher order, so that there are necessarily infinitely many different orders of real numbers, how can one speak of all real numbers at the same time?

Under the approach sketched here, we have no clear idea of *all* real numbers, but, for each given α, we can consider *all* real numbers of order α. Moreover, we can use a sort of schematic method by which we discuss all real numbers by talking about all real numbers of each order indiscriminately.

Nonetheless, it is not necessary to adopt this approach, if our purpose is just to be able to speak of infinitely many orders at the same time. Consider the system Σ_ω (or, for that matter, any Σ_α, where α is a limiting ordinal number). We can prove easily that for every bounded set x_ω of real numbers, there is a real number y_ω which is the least upper bound of x_ω. This is so, because every set x_ω is a set of real numbers of some finite order n, and its least upper bound is a real number of order $n + 1$ and *ipso facto* a set of order ω.

Therefore, as soon as we get to sets of order ω, we can already avoid the necessity of ascending to any higher orders to get least upper bounds. Similarly, the sets needed for the axiom of choice, the interpretations of the continuum hypothesis, and the Bolzano-Weierstrass theorem are also of order ω, provided each given set is of order ω (and therefore of an order n, for *some* finite n).

It follows from this remark that in order to speak of all sets of all finite orders at the same time, we need neither Russell's axiom of reducibility (see below) nor the full theory Σ which includes an indeterminate totality of orders. The system Σ_ω is more satisfactory than Russell's method which unnecessarily introduces unmanageable difficulties in connection with impredicative sets. The theory Σ is philosophically even more satisfying because it seems rather arbitrary to stop at either order ω or some other order α. For example, the fact that Σ_ω is directly subject to Gödel's theorems while the theory Σ, so far as I can see, is not, is one indication of the preferableness of Σ. We might say that the approach embodied in Σ is superior to Russell's in more than one way.

It is customary to define real numbers as certain sets of natural numbers or rational numbers. Since these sets are infinite, each of them has to be given by a law or a principle for selecting its members. Naturally, therefore, how rich a theory of real numbers is depends very much on the method of definition we are permitted to employ. We can, of course, think of many curious ways of defining sets. Consequently we do not have any clear idea of *all* sets. The notion of set and thereby that of real number is relative to our theory of definition in the sense that different theories of definition would give us different theories of sets and real numbers.

This was apparently not realized by people of the nineteenth century. They seem to talk as if there was a unique absolute theory of definition. Thus Dedekind spoke of all sets or cuts of real numbers without bothering to stop and examine what he meant by *all*. Similarly Cantor spoke of all denumerable sets in his proof of the indenumerability of the set of all real numbers.

Now, how do we know that the theory Σ does give us all the real numbers we need in ordinary mathematics?

If we examine the real numbers actually used in mathematics, we easily see that the domain has gradually been expanded. The rational numbers, of course. Then there were simple irrational numbers such as $\sqrt{2}$. The Greeks also considered more complex cases. In general, however, they seem to confine themselves to irrationalities obtained by repeatedly taking the square root, an operation performable geometrically by ruler and compass. The next logical extension is to include all algebraic numbers. There are also transcendental numbers, e.g., e and π.

Since it seems natural to think that every infinite decimal defines a real number, it is worthwhile to get an idea of the wide range of possible laws or definitions for determining these decimals. We can roughly contrast "natural" with "artificial" definitions. The natural ones arise out of the actual development of mathematics, such as those for e, π, values of certain common functions, etc. Usually there are organic connections between these and our main body of knowledge so that we have more information about the infinite decimals and real numbers determined by them. The artificial ones can be manufactured by playing with accidental characters of notation or actuality. For example, the following infinite decimal,

$$0.1223334444\ldots,$$

defined by the law that for every numeral n, n is repeated n times (for example the numeral 15 is written 15 times in succession).

Or the infinite decimal obtained from that for π by substituting an arbitrary numeral (say 7, or 891) for every digit 1 (or 2 or 3, etc.). Obviously, from each infinite decimal we can in this way generate infinitely many artificial infinite decimals about which we know nothing apart from their definitions.

A different kind of artificial definition is obtained if, for example, we determine

$$0.a_1a_2a_3a_4\ldots$$

in the following manner: beginning with January 1, 2000, a_i is 0 or 1 according to whether or not more boys are born than girls on the i-th day.

From these examples it should become clear how easy it is to make the invention of curious infinite decimals an enduring pastime. We shall, however, refrain from self-indulgence and confine ourselves to the natural definitions.

The most efficient way of generating new real numbers is by functions. Thus, given a fixed subclass D of the class of real numbers (e.g., D may be the class of rational numbers or that of algebraic numbers) and a function

$$y = f(x),$$

there is a real number y for every x in D. It may happen that for every x in D, $f(x)$ is also in D. Then the function $f(x)$ generates no new real numbers. However, it may also happen that for certain values of x in D, $f(x)$ is no longer in D. Let D′ be the class of all real numbers which are either in D or are the values of $f(x)$ for some x in D. With regard to D′, it may again happen that for some x in D′, $f(x)$ is not in D′. We can then consider a larger class D″. And so on. In general, this process can be continued indefinitely and the sum of all such classes (call it D_f) satisfies the condition that for every x in D_f, $f(x)$ is also in D_f.

In this way, given each fixed class D of real numbers and a function $f(x)$, we can try to find the corresponding D_f which in special cases may be the same as D. Hence, we can approach the totality of all real numbers in the following manner. Let us start from, say, the domain of rational numbers. Consider, e.g., the ordinary algebraic and transcendental functions. Let us add them successively and every time expand the domain of real numbers to get a larger one in which for every x in it $f(x)$ is also in it. In this way, we reach a totality of real numbers *closed* with regard to a given totality of functions.

No number of functions could determine all the real numbers. Yet we certainly want to get at least a domain of real numbers which would be *closed* with regard to all the functions we have occasion to consider in ordinary mathematics. Once we are sure that a certain theory of real numbers does provide a domain satisfying this requirement, we need not be too much concerned with the question how many more real numbers are also included.

Therefore, in order to determine the minimum requirements that a theory of real numbers is to satisfy, it is relevant to consider first the ordinary functions in analysis.

It is quite easy to prove that all real numbers which can be obtained by ordinary procedures of classical analysis can be obtained in the system Σ_ω (indeed, in a partial system, say, Σ_5).

There is presumably a great gap between the totality of all the special laws and series which we have had an opportunity to study and the totality of all possible laws and series which we may or may never get around to investigating. If so, how can we ever hope to get a satisfactory theory of all possible laws or definitions?

The answer to this rhetorical question is to distinguish two types of theories of "all" laws. If what is desired is a theory which provides us with detailed information about all the possible laws concerning some of which we may never get such information otherwise, it would almost be a tautology to say that no such theory can be obtained. On the other hand, there is no reason why a much more modest theory would not actually suffice for the purpose of setting up a rigorous foundation of real numbers and mathematical analysis. For example, we can employ a theory in which all the known laws of defining infinite decimals are included, and at the same time we deliberately use the word "all" ambiguously so that the door is open for other newcomers to join the totality of *all* laws of the theory. The theory Σ seems to be one such.

9. The axiom of reducibility. Roughly the ramified theory of types is equivalent to the system Σ_ω minus the variables of order ω. Russell's axiom of reducibility says that for every set there is a coextensional set which is of the order next above the highest order of its arguments. For example, every set whose members are objects of order 0 is, by this, coextensional with a set of order 1.

By the help of such an axiom of reducibility, statements about "all first-order functions (or sets) of m" yield most results which otherwise would require "all functions (or sets) of m". The axiom leads to all the desired results and, so far as we know, to no others. Nevertheless, Russell thinks that it is not the sort of axiom with which we can rest content and conjectures that perhaps some less objectionable axiom might give the results required. (By the way, Russell's original form of the axiom of reducibility was more complex than the currently accepted formulation, because Russell could not make up his mind exactly what methods are permitted in defining functions or sets of lowest orders. We are assuming that all and only methods of formal logic are admitted.)

The purpose of this section is to argue that there is nothing wrong in speaking of functions or sets of all orders at the same time, and to prove that we do not need the axiom of reducibility at all. Thus, for example, if we use the variables x_n, y_n, z_n, etc. ($n = 1, 2, \ldots$) to refer to sets of positive integers of order n, we seem to need either the axiom of reducibility or an infinitely long expression in order to make a statement about sets of positive integers of all orders. There is, however, nothing to prevent us from introducing in addition, as in the system Σ_ω, a new kind of variable x_ω, y_ω, z_ω, etc., which take all these sets no matter of what order, as values.

Once we introduce such general variables for all sets (of whatever order), of a same type, we can do all the things for which the axiom of reducibility was originally proposed. Consequences of the axiom which are no longer available are precisely the results that contradict the basic spirit of the constructive approach, intended by Russell's theory of ramified types.

The announced reason for introducing the axiom of reducibility was to enable us to talk about all sets or functions of certain given things in addition to all sets or functions of each given order. Let us discuss one by one how the use of general variables can substitute for the axiom of reducibility for the various situations considered by Russell.

The first thing is with regard to mathematical induction. Russell wants

https://doi.org/10.2307/2267732 Published online by Cambridge University Press

to say that a positive integer is one which possesses *all* properties possessed by 1 and by the successors of all numbers possessing them. If we confine this statement to all first-order properties x_1, we cannot infer, without using the axiom of reducibility, that it holds of second-order properties x_2. However, using general variables x_ω, y_ω, etc., for sets of positive integers of all (finite) orders, we can now make the above statement with regard to *all* properties x_ω (of positive integers).

A second use of the axiom of reducibility is with regard to the definition of identity. He defines two individuals as identical when they have the same first-order properties. By the axiom of reducibility, he then proves the theorem that two such individuals have the same properties of every order. In the system Σ_ω, we can, using general variables for all properties, adopt his theorem as the definition of identity. Thereby the axiom of reducibility is no longer needed.

A third and more important application of the axiom is in the development of the Dedekind theory of real numbers. Thus, if we define real numbers as certain sets of rational numbers satisfying Dedekind's requirements, then, since there are such sets of different orders, there are also real numbers of different orders. This, as we have observed, can again be handled in systems such as the system Σ_ω.

On the other hand, the abolition of the axiom of reducibility does entail the important consequence that Cantor's proof for the theorem that there are absolutely more real numbers than positive integers breaks down in Σ_ω, although, in a modified form, it can be gotten by the axiom of reducibility in a system with only finite orders. Indeed, it is no longer possible to prove the existence of any cardinal number greater than aleph-zero (the number of positive integers) in Σ_ω.

From the constructive point of view adopted by Russell, this is, however, not only no objection to the approach embodied in Σ_ω, but rather a point strongly in its favour. This is so, because in proving the existence of any infinity beyond aleph-zero, the impredicative definitions are indispensable, and impredicative definitions are precisely what Russell's theory set out to abolish.

In other words, the axiom of reducibility actually serves two very different purposes: (1) to enable us to speak of all sets or functions of certain things without having to enumerate the infinitely many different orders; (2) to enable us to introduce sets by impredicative definitions and properties. It seems that Russell, in trying to remedy a minor verbal difficulty, unwittingly reintroduced impredicative definitions at a back door through the axiom of reducibility (indeed, introduced an assumption which embodies the very essence of impredicative definitions in set theory). The remarks of this and the preceding sections should be sufficient to establish the conclusion that as soon as general variables are introduced, the axiom of reducibility becomes unnecessary at least for all the things which it was originally introduced to do.

In my opinion, the present approach also illuminates the criticism of Weyl by Hölder in a paper of 1926.[9] According to Hölder, since in the definition of the least upper bound of a set M of real numbers the quantification may be understood as ranging over the real numbers of M only, which are given beforehand, and not over *all* real numbers, it is not true that the definition of the least upper bound involves a circle. If we adopt the standard formalization of the least upper bound theorem where we do not admit real numbers of different orders, Hölder seems clearly wrong. But if we accept a constructive approach as in the present essay, then Hölder is right because in each given set M the real numbers must be of a definite order, and we can define a least upper bound which is of the next higher order.

I think Hölder's remarks are very interesting for the present approach because they seem to show that it is quite natural to use a ramified analysis. We might even view the systems of this essay as, among other things, attempts to formalize Hölder's interpretation of the classical theorem of least upper bound.

10. The vicious-circle principle. The theory Σ is built up in accordance with the Poincaré-Russell vicious-circle principle. Since the theory of types is also based on the same principle but differs from Σ in many ways, it is of interest to ask whether the theory Σ does not violate the principle in certain respects.

Several somewhat different forms of the principle are given by Russell, but we can confine ourselves to the one stating that no totality can contain members definable only in terms of that totality. This vicious-circle principle, according to Russell, enables us to avoid "illegitimate totalities." It follows that given an open formula (propositional function) p which either contains quantifiers referring to sets of order α or has its argument value referring to sets of order α, the set defined by p must be at least of order $\alpha + 1$.

Thus, if we start from a given totality of basic objects and call them of (say) order 0, we can proceed to define sets of orders 1, 2, 3, etc. by introducing new variables and new abstracts at successive stages. It is natural to record our advance by using such order indices. When we have gone through all finite ordinals, there is nothing to prevent us from going to transfinite orders.

Sets of order ω should be all and only those which are defined in terms of variables and abstracts of finite orders. On the one hand, since ω is

[9] Otto Hölder, *Der angebliche circulus vitiosus und die sogenannte Grundlagenkrise in der Analysis*, ***Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig***, Mathematisch-physische Klasse, vol. 78 (1926), pp. 243–250.

higher than all finite ordinals, every abstract containing only variables and abstracts of finite orders defines a set of order ω. On the other hand, since ω is the smallest infinite order, an abstract containing any variable or abstract which is not of a finite order must be of order higher than ω. Similarly, for each limiting ordinal α. As a result, for each limiting ordinal α, sets of order α serve to sum up all sets of all lower orders in the theory Σ.

There is no reason that we should stop at any particular ordinal number α of the second number class, since we can certainly proceed further and define abstracts of order $\alpha + 1$, $\alpha + 2$, etc. Hence, instead of using a definite formal system, we allow in Σ indefinitely many orders α and corresponding partial systems Σ_α.

In ***Principia mathematica***[10] it is emphasized that certain expressions are neither true nor false but meaningless. Thus, for example, it is neither true nor false to say that a set belongs to itself, because the question is meaningless. In general, '$a \in b$' is meaningless unless a and b are of suitable types (more specifically, b is just one type higher than a).

Recently, this emphasis has often got into headlines in philosophy. It is contended that many, if not all, philosophical problems arise because we want to get a 'yes or no' answer to meaningless questions. When we say that the universal set belongs to itself, or that justice is blue, we are said to be making a 'category mistake.'

However suggestive, for philosophy, the idea may be, it does not seem necessary so far as logic is concerned. An obvious way is to call the expressions false instead of meaningless. Indeed, this is followed in the theory Σ.

The procedure in ***Principia*** of treating expressions such as "$a_n \in a_n$" (n as type or level index) as meaningless rather than false leads to the consequence that the sets are divided into mutually exclusive ranges of significance. This is so because it would be extraordinarily queer and inconvenient to say, for example, that "$a_n \in a_n$" is meaningless while "$b_n \in a_n$" is meaningful (and true), or, in general, that some term can replace another term in one meaningful expression but not in another. The situation is most striking with regard to the substitution of equals for equals (substitutivity of identity). On the other hand, it is quite all right to say, for example, that "$a_n \in a_n$" is false but "$b_n \in a_n$" is true.

Hence, the fact that the ordinary theory of simple types does not permit mixture of types is closely connected with the decision of considering certain expressions as meaningless instead of false. If, for example, we introduce types or levels to the system in Σ which deals solely with sets and variables of finite orders and then add the axiom of reducibility to nullify the distinction of orders, then we get a theory which is like the simple theory of types but permits the mixing of different types.

[10] A. N. Whitehead and Bertrand Russell, Vol. I, p. 41.

It goes without saying that the use of transfinite orders (and without bounds) is the principal difference between the theory Σ and the ordinary ramified theory of types without the axiom of reducibility. The axioms of limitation provide a new feature which has often been discussed but has never been carried out before in standard forms of set theory.

It may be thought that the famous Gödel proposition, which essentially says of itself that it is not provable in a certain system, violates the vicious-circle principle. If so, it would be pretty bad for the vicious-circle principle, since Gödel's construction is a perfectly sound procedure. Actually, however, the self-reference is achieved by using considerations from outside the system. Gödel is defining by a non-objectionable method something which ordinarily can only be defined by a vicious circle. This is like proving some set defined by an impredicative definition actually equivalent to a set defined by a predicative definition, thereby making the set non-objectionable.

More exactly, if we say, "This proposition is not provable," we are using self-reference, and defining a proposition by referring to itself (or a totality including itself), but when we find a way of doing the matter as Gödel does, it is justified. We no longer define a proposition but just interpret a proposition, and prove results by means of this interpretation.

11. Predicative sets and constructive ordinals. In this essay constructive sets are identified with predicative sets. Predicative sets are sets which can be defined without violating the vicious-circle principle. It is desirable to have a more exact characterization which is, for instance, as sharp and acceptable an explication for predicativeness as recursiveness is for the intuitive concept of effective computability.

If we confine our attention to mathematical objects, one possibility is to say that a set is predicative if and only if it is coextensional with a set of the theory Σ. Leaving aside the difficult question of justifying the adequacy of the identification, we are faced also with the more urgent question of rendering the answer clear.

To determine even roughly the domain of sets available in Σ, we should have a pretty good idea of what a constructive ordinal is, since Σ is the union of all systems Σ_a where a is a constructive ordinal.

The Church-Kleene definition[11] of constructive ordinals in terms of recursive functions is clear and definite enough for the purpose. Yet it is rather narrower than what is wanted, since, in defining new ordinals, we would like to say that each monotone increasing sequence, generated by a

[11] See S. C. Kleene, *On notation for ordinal numbers*, this Journal, vol. 3 (1938), pp. 150–155, and Alonzo Church, *The constructive second number class*, ***Bulletin of the American Mathematical Society***, vol. 44 (1938), pp. 224–232.

predicative function, of constructive ordinals determines a limit which is again a constructive ordinal, and yet there are predicative functions which are not recursive. As a result, we seem to get into a circle: in order to determine the region of predicative sets, we must have first a definite notion of constructive ordinals; in order to get such a definite notion, we must determine first the region of predicative sets.

One possible way to eliminate this impasse is to begin with a definite totality of ordinals (e.g., all ordinals below the first ϵ-number, or all constructive ordinals in the sense of Church-Kleene) and then consider the totality of all ordinals β definable in some of the systems Σ_α, where α is an ordinal in the first totality. In general, the new totality contains the first totality as a proper part. We can then consider the totality of all ordinals γ definable in some system Σ_β, where β is an ordinal of the second totality. And so on. There is, of course, no obvious assurance that all constructive ordinals we want will be obtained in this fashion. Certainly a large variety of ordinals can be gotten.

The usefulness of this definition of constructive ordinals depends on the following facts: given an ordinal number α, we have, as described above, a definite procedure of constructing Σ_α; given a system Σ_β, the totality of the ordinal numbers obtainable in Σ_β is determined. No circularity is involved.

The limitations upon formal systems were discussed in § 5. Let us now make a few general observations on how the approach embodied in the theory Σ is both natural and not subject to objections raised there.

One reason why the approach is natural is the inclusion of all principles employed in standard systems. The consistency is assured by rejecting altogether impredicative sets. The possibility of immediate extension is excluded by deliberately avoiding the postulation of a highest level of sets. There is, so to speak, at every stage an indeterminate limit on our actual knowledge of the possibilities of constructing new sets from given sets. Every given determination of the limit can be transcended, but no determinate limit transcends or even exhausts all the possibilities which are permitted by the theory. In this way the second objection is evaded.

To meet the third objection, indenumerable sets are entirely excluded. Given any enumeration of sets of (say) positive integers in the theory, there is always some other set of positive integers, not included in the enumeration. Moreover, given an arbitrary order, there are sets which cannot be enumerated by sets of that order. Nevertheless, every set in the theory is denumerable by some set (relation) in the theory. There are not only no absolutely indenumerable sets, but even no sets of the theory which are not denumerable in the theory. There are only indenumerable sets relative to each order in the theory.

This meets the third objection only by adopting a certain preferred viewpoint (you might say, a philosophy) which interprets sets as in some sense constructed. From this point of view, the fact that no enumeration can exhaust all sets of positive integers is explained not by the existence of indenumerable sets but rather by the impossibility for our intellect to have a clear and distinct idea of the totality of *all* sets or laws defining enumerations, for unless we are able to contemplate such a totality, it is quite senseless to ask whether there exists any set which is indenumerable in the absolute sense. If each time we can only contemplate a portion of all sets or laws of enumeration, we can only prove that certain sets are indenumerable when we restrict our means of enumeration to the given kind.

Or, the same fact can be explained by our inability to contemplate at one and the same time the totality of *all sets* (or laws defining sets) *of positive integers*. For, it may be argued, although we cannot contemplate all laws of enumeration at the same time, we can contemplate each of them. Therefore, if we can grasp at once the totality of all sets of positive integers, we can see schematically that each law of enumeration is inadequate to an enumeration of all of these, and then conclude that the totality of all sets of positive integers is absolutely indenumerable. But if, as is natural to assert from a constructivistic view, we cannot have a clear and distinct idea of the totality of all sets of positive integers, then it is quite senseless to ask whether or not such a totality, if we could grasp it, could be exhausted by a specific law of enumeration.

From this approach, to ask whether the totality of all sets of positive integers is denumerable (in the absolute sense) is very much like asking, as a common man though perhaps not as a physicist, whether or not the world is bounded in time and space. The totality of all sets or of all sets of positive integers is like Kant's thing-in-itself, while the constructible sets correspond to all possible experience. To parrot Kant: Now if I inquire after the quantity of the totality, as to its number, it is equally impossible, as regards all my notions, to declare it indenumerable or to declare it denumerable. For neither assertion can be contained in mental construction, because construction of an indenumerable totality or a closed denumerable totality incapable of further expansion, is impossible; these are mere ideas. The number of the totality, which is determined in either way, should therefore be predicated of the transcendent totality itself apart from all constructive thinking.[12] We cannot indeed, beyond all possible construction, form a definite notion of what the transcendent totality of all sets may be. Yet we are not at liberty to abstain entirely from inquiring into it; for construction never satisfies reason fully, but in answering questions, refers us further and further back, and leaves us dissatisfied with regard to their complete solution.... The enlarging of our views in mathematics, and the possibility of new discoveries, are infinite. But limits cannot be mistaken

[12] Compare ***Prolegomena to any future metaphysics***, § 52, c.

here, for mathematics refers to the constructible only, and what cannot be an object of intuitive contemplation, such as the totality of all laws, lies entirely without its sphere, and it can never lead to them; neither does it require them.[13]

The question whether Σ can further be extended is debatable. Since we make conventions in Σ which depend on previous ones, we cannot effectively predict or well-order all possible orders used in Σ once and for all. On the other hand, it seems possible to look from outside and speak of the totality of all sets in Σ or all orders used in Σ. To speak of a universal set which contains all sets in Σ as members is either making an impossible convention or making a convention of a higher kind. It may be exaggeration to call such a convention impossible. It is just not very informative. Thus, it appears possible to add consistently to Σ an isolated universal set, provided we do not try to say too much about this transcendental set. It we wish to, we may even say that Σ is all that is needed for mathematics, introduction of sets beyond belongs to the realm of philosophy.

12. Concluding remarks. We may say that there are three regions in mathematics: (1) the effectively decidable; (2) the constructive; and (3) the transcendental. The theory Σ is an attempt to formalize the second region. Cantor's jump to the absolutely indenumerable belongs to the third region, while Brouwer's logic, as well as Hilbert's finitist viewpoint, deals mainly with the first region. Hilbert and Brouwer differ in that Brouwer never accepts willingly anything but the decidable logic, while Hilbert would allow us to use everything which can be justified on the basis of the decidable logic. This might also be expressed by saying that Brouwer requires restriction to his logic in all mathematics, while Hilbert requires it only in the domain of metamathematics.

For Hilbert, anything which can be seen or proved consistent finitistically, is acceptable: to be is to be consistent. It appears from discussions on the finitist viewpoint that the theory Σ can be proved consistent by finitist arguments. The theory Σ may perhaps be viewed as a theory satisfying Hilbert's demands for the foundations of mathematics. Of course there are finitist arguments which are not formalizable in current systems of elementary number theory.[14]

The main purpose of a construction and development of the theory will be not so much the exhibition of a formal system as the basis of all mathematics, as the presentation of an argument to justify all mathematical reasoning which does not get into the transcendental. It will be an attempt

[13] Ibid, § 57.

[14] See, e.g., D. Hilbert and P. Bernays, ***Grundlagen der Mathematik,*** vol. II, p. 373.

to show that all such reasoning can be formalized in some formal system falling under the schema Σ.

Once it is clear that all constructive mathematical reasoning, that is, all mathematical reasoning that does not involve the dubious use of the impredicative definitions or jump to the indenumerable, can be formalized in the loose framework Σ, there will be no more need to formalize each argument by making explicit the orders of the various sets concerned. Rather we can then proceed with no special attention to the orders, with the realization that whenever we wish, we can always formalize each argument in some system Σ_a. In this way, the common practice in ordinary mathematics can be justified by the theory Σ. It is the kind of justification which does not interfere with the common practice. In this sense it is again more natural than other approaches to the foundations of mathematics.

But how can we prove that this is the case? There are two methods: a long one and a short one. The long method consists in a completion of the arguments by developing in Σ the various branches of mathematics. For example, as the Bourbaki group continues to turn out more and more volumes of their treatise, we show for each volume how all the definitions and proofs can be formalized in the theory Σ. While there does not seem to be fundamental obstacles to such a program, actually to carry it out is a complex and time-consuming matter. Whether this is worthwhile is hard to decide because on the one hand, many mathematicians would undoubtedly find the result of carrying out such a program quite uninteresting, while on the other hand, without an actual carrying out of the program most logicians would suspect the soundness of the high claim.

The short method is to let the matter stand as it does in an expanded version of this essay (or, more exactly, an expanded version which includes detailed formal development of matter covered in §§ 6–8) and challenge anybody who questions the adequacy of Σ to produce some mode of inference which is used in ordinary mathematics but cannot be formalized in Σ. The trouble with this shortcut is that few mathematicians who have a more or less clear view of the whole field of mathematics are likely to care to study the theory Σ carefully.

There are many open problems concerning the relation between the decidable and the constructive, as well as the relation between the constructive and the transcendental. The theory Σ is intended to be a definitive theory of the region of the constructive so far as mathematics is concerned. It is not clear to the writer whether there is a generally accepted unique characterization of the region of the decidable. So far as the region of the transcendental is concerned, we note merely that it is of course also possible to construct a theory which is related to (say) Zermelo's full set theory in the same manner as the theory Σ is to the system Σ_0. Indeed, for the transcendentalists it may even be possible to use as order indices ordinal numbers beyond Cantor's second number class.

The most famous open problem in the field of transcendental set theory is Cantor's continuum problem or the problem whether Cantor's hypothesis is independent of the other axioms of set theory. A related problem is the independence of the axiom of choice. From the known fact that the axiom of choice is derivable from the generalized continuum hypothesis, it also follows that the independence of the axiom of choice entails the independence of the generalized continuum hypothesis. A less famous but more basic open problem is the consistency of the use of impredicative sets.

One objection to Σ is that it does not contain a maximum order so that, for example, we cannot speak of all sets, or all real numbers at the same time in Σ. This is at least partly a linguistic difficulty and can be avoided to that extent by some linguistic device. Thus, in Σ we can introduce an additional kind of general variable x, y, z, etc. so that we can assert '$(x)F(x)$' when and only when for each α, we can assert '$(x_\alpha)F(x_\alpha)$'; we can assert '$(x)(\exists y)F(x, y)$' when and only when for each α, there is a β; we can assert '$(x_\alpha)(\exists y_\beta)F(x_\alpha, y_\beta)$', etc. This device has also the additional advantage that for many purposes we can make general assertions without making the relevant order indices explicit. This draws the theory closer to the common practice in mathematics.

Since the notion of set determined by the theory Σ is more transparent than the transcendental notion of set, it seems reasonable to expect that the theory Σ may enable us to get better insight into certain mathematical problems which are difficult mainly because they are very abstract and general. For example, we may be in a better position to deal with problems which are concerned with arbitrary sets or arbitrary functions.

Gödel defends transcendental set theory by contending that it can be justified by conceiving sets and concepts as real objects and that it is legitimate so to conceive them. "It seems to me that the assumption of such objects is quite as legitimate as the assumption of physical objects 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 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."[15]

One possible interpretation of the argument is to say that it amounts

[15] KURT GÖDEL, *Russell's mathematical logic*, ***The philosophy of Bertrand Russell***, The Library of Living Philosophers, Vol. V (1944), edited by P. A. Schilpp, p. 137.

to conceding that transcendental set theory and the assumption that sets are real objects are necessary evils which we have to put up with if we want to have a fairly simple theory of mathematics. The theory Σ seems to show that the evil of accepting transcendental set theory is not necessary. In the first place, so far as the data to be accounted for are concerned, neither Gödel nor others consider it necessary to preserve Cantor's higher infinities, but nearly everybody wishes to retain classical analysis. The evil of using impredicative sets is considered necessary because it is thought that classical analysis cannot be developed without these sets. The theory Σ establishes that this is not the case.

As far as simplicity is concerned, there are of course many different senses of the word 'simplicity.' In one important sense, the demonstrable consistency of the theory Σ proves conclusively that it is simpler, at least relative to our present knowledge, than standard systems of transcendental set theory. Usually the ramified theory of types is considered to be hopelessly messy because we have to distinguish at least two hierarchies in it (the orders and the types or levels). It is, however, known that one hierarchy is enough, and the theory Σ actually uses just one hierarchy. Moreover, in formalizing actual proofs we do not have to let even the distinction of orders intrude, as long as we are careful not to use circular arguments in which impredicative definitions cannot be dispensed with. Once we have seen how much can be done in the theory Σ, we can continue to do mathematics as usual with the realization that the arguments used can be formalized in Σ if and when we wish to. Only occasionally we encounter 'strange' modes of reasoning which have to be examined more carefully before we can decide whether they are formalizable in Σ or belong to the domain of transcendental set theory. For example, most people, when confronted with Cantor's indenumerability arguments, presumably have some uneasy feeling and suspect the presence of some hidden fallacy. Undoubtedly ordinary mathematicians would consider such arguments as extraordinarily uncommon.

OXFORD UNIVERSITY