I

THE PROBLEM

The truths of mathematics hold a place of honour in our collective web of belief. While our fondest empirical beliefs—that the earth is round, that plants grow from seeds, that heavy objects fall—are all supported by definitive accumulations of good empirical evidence from observation, experiment, and theory, in the mathematical case, we reach a cherished evidential ideal: we have proofs. Granted, these proofs often rest on premisses or background assumptions, but perhaps such assumptions can themselves be proved. Indeed, they often are proved, in the literature of the field, in its canonical textbooks or its folklore; sometimes, the unproved assumptions are considered so trivial that the informed reader is supposed to be capable of sketching for herself whatever further argumentation is desired. But this process of appeal to further and further proofs can't go on for ever; eventually, we must reach fundamental assumptions.

In fact, it is these fundamental assumptions that often define a branch of mathematics, as the Peano axioms define arithmetic, or axioms for complete ordered fields define the real numbers at the basis of the calculus and higher analysis. The astounding achievement of the foundational studies of the late nineteenth and early twentieth centuries was the discovery that these fundamental assumptions could themselves be proved from a standpoint more fundamental still, that of the theory of sets. The idea is simple: the objects of any branch of classical mathematics—numbers, functions, spaces, algebraic structures—can be modelled as sets, and resulting versions of the standard theorems can be proved in set theory. So the most fundamental of the fundamental assumptions of mathematics, the only such assumptions that truly cannot be proved, are the axioms of the theory of sets itself.

In this sense, then, our much-valued mathematical knowledge rests on two supports: inexorable deductive logic, the stuff of proof, and the set theoretic axioms. The question of what grounds our faith in logical inference is a vital and profound one, but the subject of this book is the

second question: what justifies the axioms of set theory? Though it might seem that the only justificatory method of mathematics is proof, our question points to something more, to a brand of mathematical evidence that has received much less attention.

Questions of justification are always of interest to philosophers, but it should be noted that this particular question of justification is of considerable importance to contemporary set theorists as well. Though the current axioms of set theory are strong enough to found all of classical mathematics, they cannot settle some of the most basic questions of set theory itself, and as set theoretical considerations have reached into other branches of mathematics, the current axioms have also been unable to settle natural questions of algebra, topology, and so on. This situation has led to a search for additional axioms, for new fundamental assumptions, at which point the question of how such assumptions are to be justified takes on a fully practical application. My fondest hope is to provide a philosophically useful account that is also sensitive to (or better yet: relevant to) this pressing methodological concern of contemporary set theory.

My goal in Part I is to give a more complete characterization of this problem in contemporary set theoretic axiomatics. Much of this material will, of necessity, be familiar to the specialist, who is cordially invited to skip or skim. (Those most concerned with the philosophical issues surrounding realism and naturalism might prefer to do the same, returning to Part I as need or curiosity arise.) I begin (in I. 1) with a sketch of the origins of set theory and end (in I. 6) with a discussion of the particular new axiom candidate that will serve as a recurrent case study in Parts II and III. Along the way, I look more closely at the nature of set theory's foundational role (in I. 2), summarize the arguments given in defence of the standard axioms (in I. 3), describe a sampling of the independent questions left open by those axioms (in I. 4), and survey the various styles of available new axiom candidates (in I. 5).

1

The Origins of Set Theory

Set theory, as we now know it, resulted from a confluence of two distinct historical developments, one beginning from the work of Gottlob Frege from the 1870s to the early 1900s, the other beginning from the work of Georg Cantor during roughly the same period. That Frege's initial motivations were at least partly philosophical, while Cantor's were at first largely mathematical, only serves to highlight the rich conceptual roots from which the theory arose. I'll begin here with the Fregean line of development (in (i)), then turn to the Cantorian (in (ii)), and conclude (in (iii)) with a glance at the confluence.

(i) In his *Begriffsschrift* of 1879, Frege invented modern mathematical logic, but his motivating concern at that time was a project in the foundations of arithmetic: he hoped to show that Kant was wrong about arithmetic, that it is analytic, rather than synthetic, as Kant had claimed. On Kant's definitions of the relevant terms, analyticity is a feature of the content of a judgement: the concept of the predicate is contained in the concept of the subject. Frege criticized this definition as too narrow,[1] as applying only to universal affirmative statements like 'all bodies are extended', and proposed to correct this flaw, 'to state accurately what earlier writers, Kant in particular, meant by'[2] the terms analytic and synthetic; but in doing so, he shifted focus from the content of a judgement to its ultimate justification:

> This means that the question is removed from the sphere of psychology, and assigned, if the truth concerned is a mathematical one, to the sphere of mathematics. The problem becomes, in fact, that of finding the proof of the proposition, and of following it up right back to the primitive truths. If, in carrying out this process, we come only on general logical laws and on definitions, then the truth is an analytic one. (Frege (1884), § 3)

[1] See Frege (1884), § 88.
[2] See ibid. § 3.

This process of tracing the grounds for a statement of arithmetic turned out to be very difficult:

> To prevent anything intuitive from penetrating here unnoticed, I had to bend every effort to keep the chain of inferences free from gaps. In attempting to comply with this requirement in the strictest possible way I found the inadequacy of language to be an obstacle; no matter how unwieldy the expressions I was ready to accept, I was less and less able, as relations became more and more complex, to attain the precision that my purpose required. (Frege (1879), 5–6)

It was to overcome this difficulty that Frege devised the formal language of the *Begriffsschrift*.

The main philosophical and mathematical outlines of the demonstration that arithmetic is analytic appear in the *Grundlagen* of 1884. The plan is to define numbers as logical objects, and to prove the basic propositions about numbers (something like the Peano Axioms)[3] from fundamental logical laws. The logic involved, from the *Begriffsschrift*, is higher order: in addition to objects, it includes first-order concepts, under which objects do or don't fall, second-order concepts, under which first-order concepts do or don't fall, third-order concepts, under which second-order concepts do or don't fall, and so on.[4]

Frege's first step is to analyse statements of number: what is claimed, for example, when we say that there are two decks of cards on the table? Some might be tempted to say we are ascribing a property to a physical thing, but Frege points out that the very same physical thing is one deck, four suits, fifty-two cards, and ever-so-many-more molecules. A physical thing by itself, as a mere aggregate of stuff, could have any number whatsoever, if the units were appropriately specified. In Frege's parlance, what settles the unit is the concept involved; the statement that there are two decks of cards on the table is a statement about a concept—the con-

[3] The Peano Axioms (actually due to Dedekind (1888)) state that (1) every natural number has a unique successor, (2) natural numbers with the same successor are the same, (3) there is a natural number 0, which is not the successor of any natural number, and (4) mathematical induction. See Wright (1983) for an explicit derivation, and Heck (1993) for a more recent account.

[4] The more common logic today is first-order logic, which quantifies only over individuals (objects). The Peano Axioms are actually second order, because mathematical induction (if a set of natural numbers contains 0 and is closed under successor, then it is the set of all natural numbers) quantifies over sets of natural numbers (the modern counterpart to Frege's first-order concepts) as well as natural numbers themselves. The same is true of the axioms for complete ordered fields that characterize the real numbers, because they include the continuity axiom (every non-empty set of reals with an upper bound has a least upper bound).

cept 'deck on the table'—to the effect that there are distinct things x and y that fall under it and no other things fall under it. As noted, we would be inclined to say something quite different about the concept 'card on the table' or the concept 'molecule on the table'.

Given this analysis of numerical statements, Frege moves on to the notion of sameness of number: what is claimed, for example, when we say that there are as many forks on the table as there are spoons? Here Frege borrows a familiar answer from Hume: we claim that there is a one-to-one correspondence between the forks on the table and the spoons on the table.[5] Such a correspondence could be displayed by moving the forks and spoons into pairs, so that each fork is paired with a single spoon and vice versa, but the correspondence itself is a relation, a two-place concept, that links forks to spoons regardless of their location in space or time. The result of this analysis is what's now called 'Hume's Principle': for any concepts F and G, the number of Fs = the number of Gs if and only if there is a one-to-one correspondence between the things falling under F and the things falling under G.

Hume's Principle is enough to identify or distinguish numbers when they are presented to us in the form 'the number of Fs', but it will not settle more general questions like those raised in Frege's famous complaint, 'we can never—to take a crude example—decide by means of our definitions whether any concept has the number Julius Caesar belonging to it, or whether that same familiar conqueror of Gaul is a number or is not' (Frege (1884), § 56). To solve this problem, Frege requires explicit definitions of the numbers, and he proposes his well-known solution: the number of Fs = the extension of the concept 'equinumerous with F' (where two concepts are equinumerous if and only if there is a one-to-one correspondence between the things falling under them).[6] The notion of extension he does not explain, taking it to be a previously-understood notion of logic; roughly, the extension of a concept F can be understood as the collection of things falling under F. So the number of decks of

[5] That is, we claim there is a relation R between forks and spoons on the table such that (1) for any fork f, there is a unique spoon s, such that R holds of f and s (more fully: for any f, there is an s such that R holds of f and s, and for any s', if R holds of f and s', then $s = s'$), and (2) for any spoon s, there is a unique fork f, such that R holds of f and s. Notice that, despite the appearance of the word 'one' in the expression 'one-to-one correspondence', this notion is defined without presupposing any notion of number.

[6] The reader may wonder how we know that Julius Caesar is not an extension, given that we don't know he isn't a number, or why extensions are logical objects and numbers are not. For a recent statement of these and related perplexities, see Heck (1993).

cards on the table is the collection of all concepts equinumerous with the concept 'deck of cards on the table'; in our example, this would be the collection of all concepts under which two things fall (that is, under which fall an x and a distinct y and no z distinct from both of them). Then $0 =$ the collection of all concepts equinumerous with 'not equal to itself'; $1 =$ the collection of all concepts equinumerous with 'equal to 0'; $2 =$ the collection of all concepts equinumerous with 'equal to 0 or 1'; and so on.

In this way, numbers are defined as logical objects. The next step is to use this definition, along with the fundamental principles of logic, to prove the basic propositions of arithmetic. This Frege did, sketchily in the *Grundlagen*, in detail in the *Grundgesetze* (vol. i in 1893; vol. ii in 1903). One among the logical laws from which arithmetic is there derived is the crucial Basic Law V: for all concepts F and G, the extension of F equals the extension of G if and only if the same things fall under F and G, or, in symbols, $\hat{x}Fx = \hat{x}Gx$ iff[7] $\forall x(Fx \equiv Gx)$. In the preface to the first volume of the *Grundgesetze*, Frege expressed some discomfort over this assumption:

> A dispute can arise, so far as I can see, only with regard to my Basic Law . . . V, which logicians perhaps have not yet expressly enunciated, and yet is what people have in mind, for example, where they speak of the extensions of concepts. I hold that it is a law of pure logic. In any event the place is pointed out where the decision must be made. (Frege (1893), 3–4)

This old worry recurred to Frege some years later, in 1902, when vol. ii was in press and he received an admiring letter from a British student of his work, Bertrand Russell. Along the way, Russell remarks that 'there is just one point where I have encountered a difficulty' ((1902), 124), and he goes on to derive what is now known as Russell's paradox.

The proof is easy.[8] First, define a new relation, membership, symbolized '$\in$' as follows: y is a member of z iff z is the extension of some concept that y falls under, or, in symbols, $y \in z$ iff $\exists F(z = \hat{x}Fx \wedge Fy)$. Then it is easy to show that for any y, $y \in \hat{x}Fx$ iff Fy.[9] Now consider $x \in x$. This says that x is the extension of a concept under which x falls; an ex-

[7] 'Iff' abbreviates 'if and only if'.

[8] Perhaps not quite so easy as the familiar version in set theoretic (rather than Fregean) notation: consider $R = \{x \mid x \notin x\}$. Then $R \in R$ iff $R \notin R$.

[9] If $y \in \hat{x}Fx$, then by definition of '$\in$', there is a concept G such that $\hat{x}Fx = \hat{x}Gx$ and Gy. Then by Basic Law V, $\forall x(Fx \equiv Gx)$, so in particular, $Fy \equiv Gy$, and thus, Fy. Conversely, if Fy, then $\hat{x}Fx = \hat{x}Fx$ and Fy, so by definition of '$\in$', $y \in \hat{x}Fx$.

ample would be the extension of the concept 'z is infinite', because (presumably) there are infinitely many infinite extensions (so that $\hat{z}$(z is infinite) is the extension of a concept—'z is infinite'—under which $\hat{z}$(z is infinite) falls). On the other hand, consider $x \notin x$. This says that x is the extension of a concept under which x does not fall; an example would be the extension of 'z is red', because (presumably) no extension is coloured at all, let alone coloured red. So $\hat{x}(x \notin x)$ contains some extensions and not others. The paradox arises when we ask if it contains itself. If $\hat{x}(x \notin x) \in \hat{x}(x \notin x)$, then $\hat{x}(x \notin x) \notin \hat{x}(x \notin x)$, and vice versa.

Frege was deeply shaken by Russell's letter:

> Your discovery of the contradiction caused me the greatest surprise and, I would almost say, consternation, since it has shaken the basis on which I intended to build arithmetic. . . . with the loss of my [Basic Law] V, not only the foundations of my arithmetic, but also the sole possible foundations of arithmetic, seem to vanish. (Frege (1902), 127–8)

It is a tribute to the man that he was able to add, 'In any case, your discovery is very remarkable and will perhaps result in a great advance in logic, unwelcome as it may seem at first glance' (ibid. 128). It seems that Russell, whose fascination with antinomies traced back to his studies of Kant and Hegel, became convinced of the lasting importance of his latest paradox only when faced with Frege's profound dismay.[10] Frege himself rushed to add an appendix to vol. ii of the *Grundgesetze* that begins: 'Hardly anything more unfortunate can befall a scientific writer than to have one of the foundations of his edifice shaken after the work is finished' (Frege (1903), 234). He goes on to suggest a hurried amendment to the system, but this was later shown to be inconsistent, assuming there are at least two things.[11] Toward the end of his life, Frege gave up the idea of founding arithmetic on logic and sought to base it on geometry instead.[12]

Recent commentators have noted that Frege's versions of the basic propositions of arithmetic can be derived from Hume's Principle alone,[13] that the fatal Law V is only needed to derive Hume's Principle itself from the definition of number. Still more recent commentators have argued that Frege himself was aware of this fact, that Frege himself provided

[10] See Moore (1988) for a history of Russell's efforts.

[11] For discussion, see Quine (1955a) and Geach (1956).

[12] See Frege (1924/5).

[13] Heck (1993) credits Parsons (1965) with the first observation that arithmetic follows from Hume's Principle; actual derivations appear in Wright (1983).

the derivation of arithmetic from Hume's Principle, a result aptly dubbed 'Frege's Theorem'.[14] These observations raise a wealth of fascinating and important questions: why did Frege find it unacceptable to adopt Hume's Principle in place of Basic Law V as a 'primitive truth'? What is the real force of the Julius Caesar problem? Why do extensions, but not numbers, count as logical objects? These questions and their close relatives form a vital part of contemporary Fregean scholarship.

Historically, however, the failure of Frege's project led to a careful examination of Russell's paradox, in the hope that it might be circumvented. And Russell's wasn't the only paradox up for consideration at the time. An example with a different flavour is Berry's paradox: as there are only finitely many descriptions of numbers in fewer than eighteen syllables, there must be a least number not nameable in fewer than eighteen syllables; but this number has just been named in seventeen. Russell himself progressed through several diagnoses of the root cause of these and other paradoxes, but he finally settled on the notion of a vicious circle.[15] Such circles are forbidden by the Vicious Circle Principle (VCP) on which the system of *Principia Mathematica* is based. In this daunting three-volume work (published between 1910 and 1913), Russell and Whitehead set out to provide a foundation for all of classical mathematics.

The VCP, as stated in the introduction to the *Principia*, takes two forms: the first rules out any collection with members that 'involve' or 'presuppose' the collection as a whole; the second rules out those with members that 'can only be defined by means of' the collection as a whole.[16] So, for example, 'the least number nameable in fewer than eighteen syllables' presupposes a collection of names, but that collection contains a name—'the least number nameable in fewer than eighteen syllables'—that can only be defined in terms of the collection. According to the VCP, such a collection of names cannot exist.[17] Another way to phrase this restriction is to forbid an object to be defined by making reference to a collection of which that object is a member. Such definitions are called 'impredicative'. Consider, for example, Russell's paradox, stated this time in more generic terms: we define Russell's collection—'the collection of all collections that are not members of themselves'—

[14] See Burgess (1984), Hodes (1984), Boolos (1987; 1990), and Heck (1993).
[15] A similar diagnosis appears in Poincaré (1913), chs. IV and V.
[16] See Russell and Whitehead (1910), 37.
[17] See ibid. 63–4.

then ask if it is a member of itself. But the very definition of Russell's collection makes reference to 'all collections', a collection to which Russell's collection itself is supposed to belong.[18] So the paradox rests on an impredicative definition, which is ruled out by the VCP.

These elementary applications of the VCP depend only on the second version—that a collection cannot contain members definable only in terms of that collection—but the development of the full logical system of *Principia* requires appeal to the first version as well, plus some auxiliary assumptions about the nature of propositional functions and quantifiers.[19] To see how this works, consider the proposition 'Socrates is wise'. A propositional function is generated when one of the terms of the proposition is replaced by a variable, as in '*x* is wise' or 'Socrates is *x*'.[20] Russell holds that a propositional function presupposes or involves its values, that is, that '*x* is wise' presupposes or involves 'Socrates is wise', 'Plato is wise', 'Trollope is wise', and so on. From this auxiliary supposition, and the VCP in its first form, it follows that a propositional function cannot appear in one of its own values; if it did, the collection of its values would include a member that involves or presupposes the entire collection. Thus, '"*x* is wise" is wise', 'the extension of "*x* is wise" is wise' and the like, must all be rejected as meaningless.[21] (Notice that '*x* is wise' is not defined in terms of the collection of its values, so it isn't the second version of the VCP that's being used here.)

The second auxiliary assumption concerns quantification. This time, consider the claim 'Plato has all the virtues of Socrates'. From this and 'Socrates is wise', it should follow that 'Plato is wise'. The quantifier 'all' seems to range over all properties of individuals; among these, any that are virtues possessed by Socrates must also be possessed by Plato. But notice, '*x* has all the virtues of Socrates' is itself a property of individuals; in particular, it is a property of individuals that refers to all properties of individuals. If we add the assumption that a proposition with a

[18] The account of Russell's paradox given ibid. 62–3, is more subtle than this one, depending, as it does, on an additional doctrine of the *Principia*, the no-class theory (see III. 2).

[19] See Gödel (1944) for a penetrating and influential study of Russell's views.

[20] If a proposition is some sort of metaphysical existent, rather than a sentence, it makes little sense to speak of substituting a variable into it. In fact, an ambiguity between these two readings runs through the *Principia*, to dire effect. For discussion, see Quine (1941) or (1967). (Metaphysics is the study of what things are like in themselves, independently of our knowledge or cognition; in contrast, epistemology is the theory of knowledge.)

[21] See Russell and Whitehead (1910), 39.

quantifier involves or presupposes everything in the range of that quantifier—this is Russell's second auxiliary assumption—then 'x has all the virtues of Socrates' is a property of individuals that involves or presupposes all properties of individuals; in other words, it is a property of individuals that involves or presupposes a collection to which it itself belongs. This is ruled out by the first version of the VCP. If 'x has all the virtues of Socrates' is to be meaningful, the 'all' cannot range over all properties of individuals; the range of 'all' must be restricted in such a way as to exclude 'x has all the virtues of Socrates'.

To accommodate these two restrictions—on the values of propositional functions and on the ranges of their quantifiers—Russell and Whitehead devised the theory of ramified types. The picture comes in three steps. Begin with a simple hierarchy of propositional functions without any quantifiers: (0) individuals, like Socrates; (1) quantifier-free propositional functions of individuals, like 'x is wise'; (2) quantifier-free propositional functions of things in (1) and possibly (0), like 'Socrates is x' or 'y is x'; (3) quantifier-free propositional functions of things in (2) and possibly (1) and (0); and so on. We now allow the variables in these quantifier-free propositional functions to be quantified. Order 0 consists of individuals; order 1 consists of propositional functions generated by adding quantifiers to propositional functions in (1) above; order 2 consists of propositional functions generated by adding quantifiers to propositions functions in (2) above, and so on. Within these orders, we distinguish types,[22] depending on the nature of the quantifications, so e.g. 'y has all the virtues of Socrates' and 'if x is a virtue of Socrates, then everybody has x' are both of order 2, because they are both generated from 'if x is a virtue of Socrates, then y has x', which is in (2) above, but they are of different types within that order, because one uses quantification over items of order 1 (virtues of persons) and the other uses quantification over items of order 0 (persons).[23]

Consider now the propositional functions that take individuals as arguments. Some of these—like 'x is wise'—are of order 1, the smallest order compatible with their having items of order 0 as arguments. But others—like 'x has all the virtues of Socrates'—are of order 2, because they use quantifiers over items of order 1. Propositional functions of the

[22] The reader is warned that Russell's use of the terms 'order' and 'type' do not coincide with the now-standard uses in simple type theory, discussed below.

[23] Strictly speaking, quantification is only over a type, rather than a full order, so even the description given here is too loose, but the finer details are not crucial for this overview.

first sort, those of the smallest order compatible with their arguments, Russell calls 'predicative'; the others he calls 'impredicative'. The connection with the non-technical use of these terms should be obvious: an impredicative propositional function Fx of individuals uses a quantifier over (some) propositional functions of individuals (though the VCP requires that the quantifier range over a type of propositional functions of individuals to which Fx itself does not belong).

The theory of ramified types is obviously quite complex; in fact, most commentators would agree that Russell's presentation is so imprecise as to leave some crucial matters obscure.[24] But there can be no doubt that the ramification inspired by the VCP ultimately undermines the derivation of classical mathematics in the system. This can be seen in a number of ways, the most dramatic of which concerns a fundamental theorem of analysis:[25] the real numbers are complete, that is, whenever the reals are divided into two disjoint,[26] non-empty sets A and B, with each member of A smaller than each member of B, then there is a real number r that is the greatest member of A or the least member of B. The trouble with the proof is that r is defined in terms of 'all reals'; in ramified type theory, if A and B are of order n, then r is of order $n + 1$, and the reals of order n have not been proved complete. Without this theorem, classical analysis is crippled.[27]

To circumvent this difficulty, Russell and Whitehead introduce the Axiom of Reducibility, which states that every propositional function is extensionally equivalent[28] to some predicative proposition function. To

[24] For example, Gödel writes, '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 . . . that it presents in this respect a considerable step backward as compared with Frege' ((1944), 120). A particularly serious defect is the use/mention confusion diagnosed by Quine (1941; 1967).

[25] In fact, the difficulties in deriving mathematics from the theory of ramified types actually arise before analysis, in the theory of natural numbers. The standard Fregean definition of 'natural number' is 'x is a natural number iff x has every inductive property that 0 has' (where a property P is inductive if and only if the successor of n has P whenever n does). From this, mathematical induction is derived: if 0 has property P and P is inductive, then every natural number has P. But if the properties quantified over must be restricted to a single order, it isn't clear that all the usual derivations go through. In fact, Myhill has shown that they do not; see Parsons (1990), 111, for discussion and references.

[26] That is, A and B have no elements in common.

[27] The study of what can be obtained under predicative restrictions, without the Axiom of Reducibility, begins with Weyl (1918). For a more recent discussion, see Feferman (1964).

[28] Fx and Gx are extensionally equivalent iff they are true of exactly the same arguments.

see how this helps, suppose we define the real numbers for the first time at order n, and set out to show that these reals are complete. We suppose they are divided into sets A and B, as above, and we define our r of order $n + 1$. But r is of an order greater than its argument requires—it's a real like those of order n, so its arguments are of the same type as theirs, yet it contains a quantifier over reals of order n—that is, it is impredicative. The Axiom of Reducibility then provides a predicative real r', extensionally equivalent to r, which is to say, a real r' of order n that acts exactly like r. So the reals of order n are complete after all.

Though the Axiom of Reducibility allows the derivation of classical real number theory, and the rest of classical mathematics, to go through, ramified type theory with the Axiom cannot ultimately be accepted. The trouble is that the Axiom undermines the motivation behind the ramification of types; in Quine's words,

> If for every propositional function there is a coextensive predicative one, then the symbols for propositional functions could have been construed from the start as referring outright just to the corresponding predicative ones. In short, the types of propositional functions could have been described in the first place as depending simply on the types of the arguments. The axiom of reducibility is self-effacing: if it is true, the ramification it is meant to cope with was pointless to begin with. (Quine (1967), 152)

Furthermore, the very success of the system of ramified types with the Axiom of Reducibility, that is, its success in implying a classical theory of real numbers, shows that it does not satisfy the VCP in its second form; in Gödel's words,

> It is demonstrable that the formalism of classical mathematics does not satisfy the vicious circle principle in its [second] 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* . . . does not satisfy the vicious circle principle in its [second] form. (Gödel (1944), 127)

In sum, the needs of classical mathematics require the addition of the Axiom of Reducibility to the theory of ramified types, which essentially collapses the ramification and produces a theory of simple types (that is, one in which the type of a propositional function depends only on the types of its arguments). And the VCP in its strongest form is thereby jettisoned.

Does this mean that the paradoxes are reinstated? The answer seems

to be yes and no. Paradoxes like Russell's are avoided by simple type theory: because a propositional function is of a higher type than its arguments, it makes no sense to ask whether it applies to itself, but Russell's paradox requires just this; when we ask if a set is self-membered, which we do in defining the Russell set, we are, in the context of the *Principia*, asking if the propositional function that defines the set applies to itself. Matters are not so simple with paradoxes like our second example, that of the smallest number not definable in fewer than eighteen syllables. Here the notion of 'definability', with its implied quantifier over 'all definitions', plays a key role, and simple type theory alone promises no cure.

At this point, commentators make two crucial points. First, there is Ramsey's observation that paradoxes like Russell's are those 'which, were no provision made against them, would occur in a logical or mathematical system itself. They involve only logical or mathematical terms such as class and number, and show that there must be something wrong with our logic or mathematics' (Ramsey (1925), 20). Paradoxes like Berry's, on the other hand, 'are not purely logical, and cannot be stated in logical terms alone; for they all contain some reference to thought, language, or symbolism, which are not formal but empirical terms. So they may be due not to faulty logic or mathematics, but to faulty ideas concerning thought and language' (ibid. 20–1).[29] Given that our goal is to found of mathematics, paradoxes of the second variety can be safely ignored. The second observation, more tangential to our purposes, is that the shortcomings of the ramified theory render its purported solutions to the Berry-like paradoxes problematic, anyway.[30] Under the circumstances, it seems reasonable to leave the ramified theory and concentrate instead on the simple theory of types.[31]

On the simple theory, type 0 consists of individuals, type 1 of properties (or collections) of individuals, type 2 of properties (or collections) of items in type 1, and so on. Here, as in the ramified theory, two additional assumptions are needed:[32] the Axiom of Infinity (there are infinitely many individuals) and the Axiom of Choice (to be considered in

[29] Ramsey credits Peano with a similar analysis of the second group of paradoxes.

[30] See Quine (1967), 152.

[31] In Quine's memorable phrase, 'The whole ramification, with the axiom of reducibility, calls simply for amputation' ((1941), 25).

[32] Notice that the presence of these additional assumptions, in any version of type theory, raises the question of whether or not the system is properly characterized as purely logical, and hence the question of whether or not mathematics is being founded on pure logic, as Frege originally intended.

some detail later) for objects of each type. On the basis of this theory, classical mathematics can be derived, though with some infelicities and inconveniences.[33] Even these can be eliminated if mixed types are allowed, e.g. if collections of type 2 are allowed to include items of type 0 as well as items of type 1. The theory with mixed types is called the theory of cumulative types; are there good reasons to prefer simple type theory, despite the advantages of cumulative type theory?

Notice, first of all, that the VCP alone[34] does not adjudicate between the two: in both simple and cumulative type theory, an item of type n collects only items of earlier type, so there is no danger of vicious circles. In the *Principia*, Russell presents another, essentially Fregean, argument for the simple, as opposed to cumulative, aspect of his theory. For Frege, a concept (or function) differs from an object in being 'unsaturated', that is, it stands in need of supplementation. For example, if it is a first-level concept, it needs supplementation by an object, as Socrates supplements 'x is wise' to form the 'saturated' proposition 'Socrates is wise'. A higher-level concept, like 'Plato and Socrates have P in common' is also unsaturated, but it needs a first-level concept, like 'x is wise', with an appropriate unsaturatedness of its own, to become saturated, as in 'Plato and Socrates have wisdom in common' (or 'Plato and Socrates both satisfy "x is wise"'). It wouldn't make sense to substitute an individual, that is, an item of type 0, for x in 'Plato and Socrates have x in common', because the x place in this concept is designed for a first-level concept, and the individual would not have the appropriate sort of unsaturatedness.[35] In Gödel's analysis, 'a function cannot replace an individual in a proposition, because the latter has no ambiguity to be removed, and . . . functions with different kinds of arguments (i.e. different ambiguities) cannot replace each other; which is the essence of the theory of simple types' (Gödel (1944), 136–7). This is often called the 'direction inspection' argument for the simple theory of types.[36]

[33] See Fraenkel, Levy, and Bar-Hillel (1973), 158–61, 191, for a concise discussion.

[34] Meaning, of course, the first version, the only one that remains viable when we insist on a classical theory of real numbers.

[35] Formally, let 'x is wise' be Wx and 's' stand for Socrates. Then 'Socrates is wise' becomes Ws, and the unsaturatedness of 'x is wise' has been filled by 's'. Furthermore, if 'Plato and Socrates have P in common' is symbolized $Pp \wedge Ps$, then the unsaturatedness of the variable P can be filled by 'Wx' to form $Wp \wedge Ws$. But if we attempt to substitute 's' for 'P', we get $sp \wedge ss$, which makes no sense at all, since neither s nor p has an unsaturated place that allows it to combine with itself or the other.

[36] See Russell and Whitehead (1910), 47–8.

Unless we buy the direction inspection argument, we are free to embrace the cumulative theory. I leave the matter here for now. Both simple type theory and cumulative type theory provide adequate foundations for classical mathematics, though cumulative type theory is the more workable of the two.

(ii) The second line of development contributing to the modern theory of sets stems from Cantor's mathematical work in the 1870s.[37] Cantor's initial goal was to describe the behaviour of infinite trigonometric series in terms of their sets of 'exceptional points'. He began by showing that a function given by a trigonometric series that converges at every point is uniquely represented by that series; he then generalized this result to series that converge at all but a finite number of points—these are the 'exceptional points'—then to series that converge at all but an infinite number of points, as long as those points accumulate at a single point,[38] and from there to series with more and more complicated sets of exceptional points. This detailed analysis required a precise account of the real numbers—rational and irrational—which Cantor also provided, more or less concurrently with various equivalent formulations, most notably Dedekind's.[39]

The problem solved by Cantor's theorem on trigonometric series and the pattern of its generalizations all followed a familiar line of mathematical development, growing out of natural questions in mathematical analysis, but Cantor soon pushed the enquiry in ground-breaking directions of his own. To see how, consider the elaboration of those sets of exceptional points: first an infinite set with one accumulation point; then an infinite set with finitely many accumulation points; then an infinite set with infinitely many accumulation points, but whose set of accumulation points has only one accumulation point; and so on. This thinking leads naturally to the notion of a 'derived set': the derived set of A is the set of its accumulation points. In general, if A is a point set, let

$$A_0 = A$$

[37] For a full discussion of the development of Cantor's thought, see Dauben (1979).

[38] That is, there is a single point x with exceptional points arbitrarily close to it.

[39] Cantor understood the real numbers in terms of Cauchy sequences of rationals; Dedekind used 'cuts' in the rationals instead. (See Enderton (1977), ch. 5, for a readable exposition.) Both treatments involved regarding infinite items (sequences or sets) as completed and subject to further manipulation, which brings the completed infinite into mathematics unambiguously for the first time (see I. 3. v).

A_1 = the derived set of A_0

.

.

.

A_{n+1} = the derived set of A_n

.

.

.

If the original set A is rich enough (e.g. if it is the set of reals itself), then all the A_ns will be non-empty, and there is room for

A_ω = the intersection of the A_ns

We can then form

$A_{\omega+1}$ = the derived set of A_ω

and so on.

Two separate themes are emerging here, both of which inspired Cantor to bold conceptual innovations. The first begins from the realization that infinite sets of reals can be quite subtle and complex, but nevertheless be 'small' enough not to disrupt the uniqueness theorem. This thought led Cantor to ponder the relationships between these sets of reals and the natural numbers, and between these sets of reals and the entire set of reals; the breakthrough was to think in terms of cardinality comparisons. So, for example, Cantor asked: are there more real numbers than natural numbers? He had already shown that, despite appearances, there are no more rational numbers than natural numbers, in the sense that the rationals can be put into one-to-one correspondence with the naturals,[40] but the answer to this further question was not immediately forthcoming. He first proved the answer to be yes—there are more reals than naturals—in 1873, and of all Cantor's accomplishments, this—along with the later (1891) generalization that the power set[41] of any set is larger than that set—is the one now dubbed 'Cantor's Theorem'.[42] The stunning discovery that infinity comes in different degrees led to the

[40] See Enderton (1977), 130.

[41] The power set of a set A, written $\wp(A)$, is the set of its subsets; e.g. the power set of the set of natural numbers is the set of all sets of natural numbers.

[42] See Enderton (1977), 132.

theory of infinite cardinal numbers,[43] the heart of contemporary set theory.

The second theme Cantor gleaned from his early work on derived sets came some years later, when he focused not on the sets of reals involved, but on the subscripts used in the definition, especially the infinite ordinal numbers:[44] ω, $\omega + 1$, $\omega + 2$, and so on.[45] In cardinal terms, these are all 'countable'—that is, capable of being brought into one-to-one correspondence with ω itself—but the set of all such countable ordinals, the next ordinal after all the countable ordinals, is itself uncountable.[46] This process leads to ever-increasing infinite ordinals, each of the next highest cardinality: $\aleph_0$ = the cardinality of ω, the set of all finite ordinals; $\aleph_1$ = the cardinality of ω_1, the set of all countable ordinals; $\aleph_2$ = the cardinality of ω_2, the set of all ordinals of size $\aleph_1$; and so on.[47] Cantor's first innovation was to treat cardinality as strictly a matter of one-to-one correspondence, so that the question of whether two infinite sets are or aren't of the same size suddenly makes sense; his second innovation was to extend the sequence of ordinal numbers into the transfinite, forming a handy scale for measuring infinite cardinalities. And all this began as a study of point sets, the study of ordinary sets of real numbers.

For the record, we should note that despite Cantor's boldness in bringing the transfinite within the purview of mathematical treatment, he drew the line at another kind of infinity, which he called the Absolute. By the mid-1890s, Cantor was aware that there could be no set of all sets, as its cardinal number would have to be the largest cardinal number, while his own theorem shows that for any cardinal there is a larger. Similarly, there could be no set of all ordinals, because it would then itself be the largest ordinal, and hence a member of itself, and hence smaller than itself. Nor could there be a set of all cardinal numbers, as this set would be equinumerous with the (non-existent) set of all ordinals. But Cantor was not disturbed by these results; rather, he took them to show that, in addition to the mathematically accessible transfinites, there is also an

[43] Cardinal numbers tell 'how many'—one, two, three . . . —as opposed to ordinal numbers, which tell 'how many-ith'—first, second, third, . . .

[44] See previous note.

[45] Following contemporary practice, think of ω as the set of all natural numbers (that is, the set of all finite ordinals).

[46] That is, not countable. For example, Cantor's Theorem says that the real numbers are uncountable.

[47] In contemporary set theory, a cardinal number is identified with the first ordinal of its size, so that $\aleph_0 = \omega$, $\aleph_1 = \omega_1$, $\aleph_2 = \omega_2$, etc.

Absolute Infinite, which cannot be treated mathematically. This Absolute Infinite Cantor associated with God. The merely transfinite is limited—for each there is a larger—and hence within the reach of mathematical methods; these transfinite numbers lead directly to the Absolute Infinite, which is limitless and beyond human understanding.[48]

But to return to Cantor's mathematical (as opposed to theological) development, the leading question concerned the cardinality of the continuum. Fairly simple cardinal arithmetic established that it is $2^{\aleph_0}$, and Cantor's Theorem establishes that it is greater than the cardinality of ω, that is, greater than $\aleph_0$. Cantor believed that it was, in fact, $\aleph_1$—this is the famous Continuum Hypothesis[49]—but he couldn't prove it. In fact, he couldn't even prove that $2^{\aleph_0}$ must be one of the alephs (that is, one of $\aleph_0, \aleph_1, \aleph_2, \ldots$). To show this, he needed to show that the reals can be well-ordered;[50] in fact, he believed that every set can be well-ordered, but again, this was something he could not prove.

The Well-Ordering Theorem was eventually proved, not by Cantor, but by Zermelo, in his (1904). The proof bases the theorem squarely on a new fundamental assumption, now known as the Axiom of Choice: if $\mathfrak{I}$ is a family of disjoint, non-empty sets, then there is a set C (the choice set for $\mathfrak{I}$) that contains exactly one element of each set in $\mathfrak{I}$.[51] Various forms of this axiom had been used unconsciously by many early set theorists, but Zermelo's use of the full axiom to prove a controversial result provoked a barrage of criticism from the international mathematical community.[52] In response, Zermelo published two further articles: one

[48] Dauben (1979) tells the fascinating story of Cantor's dealings with the Catholic theologians. The concrete existence of the transfinites seemed to bring God (the infinite) into the world, and thus threatened to lead to pantheism, but Cantor's delineation of the Absolute infinite, which is not in this world, defused this doctrinal worry. Constantin Guberlet went so far as to use Cantor's theory as the basis for a Berkeley-style argument for the existence of God: the entire sequence of transfinite numbers exists in the mind of God (see Dauben (1979), 143).

[49] The Generalized Continuum Hypothesis says that for all ordinals α, $2^{\aleph_\alpha} = \aleph_\alpha$.

[50] A well-ordering of a set A is a relation, R, on A that is reflexive (for all $a \in A$, Raa), transitive (for all $a,b,c \in A$, if Rab and Rbc, then Rac), antisymmetric (for all $a,b \in A$, if Rab and Rba, then $a = b$), and well-founded (for any $A' \subseteq A$, if $A' \neq \emptyset$, then there is an R-least element of A, i.e. there is an $a \in A'$ such that for all $b \in A'$, Rab). Intuitively, after any initial segment of such an ordering, there is a 'next' element, namely, the R-least member of the remaining elements.

[51] As noted earlier, a well-ordering requires that there always be a 'next' element among those remaining; the axiom of choice can be used to make such a selection. For a proof of the Well-Ordering Theorem along these lines, see e.g. Levy (1979), 160–1.

[52] For some discussion, see I. 1. 3, below. Moore (1982) gives a wonderful history of this controversy and its resolution.

contains a more careful and explicit proof of the Well-Ordering Theorem (Zermelo (1908a)); the other presents the first axiomatization of set theory and develops the theory of cardinal numbers through Cantor's Theorem on the basis of those axioms (Zermelo (1908b)). Zermelo's goal in both these papers is to put set theory in general, and his proof of the Well-Ordering Theorem in particular, on a sound foundation. To the extent that the paradoxes jeopardize this goal, he is concerned with the paradoxes[53]—he pauses, for example, to show that the usual derivations of the paradoxes are not possible in his system—but he concentrates his attention on the project of selecting a simple, efficient, and powerful set of axioms from among the jumble of controversial and mutually exclusive set theoretic principles being debated at the time. In this effort, he is guided, quite pragmatically, by the needs of his proof, and more broadly, by the need to reproduce the central theorems of the theory of sets. His list of axioms, with some adjustments and additions, forms the bulk of the Zermelo–Fraenkel set theory (ZFC) we use today (see I. 3).

This, then, was the source of the first explicit axiomatization of set theory. Its goals were practical—find a short, simple list of acceptable set theoretic statements which will allow the derivation of all important results of informal set theory, but will not allow the derivation of the known paradoxes—and its approach was pragmatic. Some years later, presenting a second version of his axiomatization, Zermelo took a deeper, more theoretical look at the structure of the world of sets.[54] This time, inspired by a new axiom, the Axiom of Foundation (see I. 3. ix), Zermelo described the models of his theory in a systematic way, as arranged in the series of stages: the first stage, V_0, is the set of all non-sets, that is, the set of ordinary objects, which Zermelo calls 'urelements'; the second stage, V_1, includes all objects and all sets of objects, in other words, $V_0 \cup \wp(V_0)$,[55] where $\wp(V_0)$ is the power set of the first stage; the third stage consists of objects, sets of objects, and sets of objects and sets of objects, in other words, $V_1 \cup \wp(V_1)$, and so on. In general, for any ordinal α, $V_{\alpha+1}$ is $V_\alpha \cup \wp(V_\alpha)$, and if λ is a limit ordinal,[56] then V_λ is the union of

[53] Actually, Zermelo discovered Russell's paradox two years before Russell did, but he did not consider it a serious threat to the theory of sets. See Moore (1982), 89, 159.

[54] See Zermelo (1930).

[55] $V_0 \cup \wp(V_0)$ is the union of V_0 and $\wp(V_0)$; the union of two sets A and B is the set whose members are the members of A and the members of B.

[56] A limit ordinal is an ordinal with no immediate predecessor, like ω.

the V_αs for $\alpha < \lambda$. This picture of sets as arrayed in a series of stages is often called 'the cumulative hierarchy'.

(iii) Now recall that the simple theory of types begins with a type of individuals, presumably the same things as the ordinary objects of Zermelo's V_0. If the vaguely described 'collections' of type 1 are taken to be sets, then type 1 consists of $\wp(V_0)$, type 2 of $\wp(\wp(V_0))$, and so on. The only difference, for example, between type 2 and V_2 is that V_2 includes both V_1 and V_0. In other words, Zermelo's cumulative hierarchy is essentially what we previously described in the theory of cumulative types; and even in the context of type theory, the cumulative theory is preferable.[57] This, then, is the confluence of the two sources of modern set theory; both lines of thought lead us to the idea of cumulative stages, which is brought to mathematical precision in the cumulative hierarchy.[58]

Finally, one more point of convergence. We've seen that Frege's initial motivation was philosophical while Cantor's was mathematical; Frege wanted to show that arithmetic is analytic while Cantor wanted to prove theorems about trigonometric series. But it must be admitted that Frege's motivations included a mathematical one: he hoped to complete the mathematical project of rigorization that began with Berkeley's criticisms of the calculus, continued through the work of Cauchy and Weierstrass on continuity and convergence, and led up to Cantor and Dedekind's accounts of real numbers.[59] Frege writes:

> The concepts of function, of continuity, of limit and of infinity have been shown to stand in need of sharper definition. Negative and irrational numbers, which had long since been admitted into science, have had to submit to a closer scrutiny of their credentials. . . . In all directions these same ideals can be seen at work—rigour of proof, precise delimitation of extent of validity, and as a means to this, sharp definition of concepts. . . . Proceeding along these lines, we are bound eventually to come to the concept of Number and to . . . the foundation of the whole of arithmetic. (Frege (1884), §§ 1–2)

Frege's goal was to show that those foundations lay in pure logic, and Russell continued on this path.

Meanwhile, Cantor's narrowly mathematical goal of proving theo-

57 Assuming we remain unmoved by the direct inspection argument.
58 Cf. Gödel (1933), 45–9.
59 See Demopoulos (1994) for discussion.

rems about trigonometric series was soon overshadowed by the larger project as he moved into the general theory of sets. By 1884, the very year Frege broached the foundations of arithmetic in his *Grundlagen*, we find Cantor writing: 'pure mathematics . . . according to my conception is nothing other than pure set theory' (Cantor (1884a), 84). Here set theory, not pure logic, is cast as the ultimate foundation for mathematics, but given the eventual convergence of ontologies noted above,[60] this difference seems less significant. This perspective continues in Zermelo's thinking: 'Set theory is that branch of mathematics whose task is to investigate mathematically the fundamental notions "number", "order", and "function", taking them in their pristine, simple form, and to develop thereby the logical foundations of all of arithmetic and analysis' (Zermelo (1908b), 200). So the two strands of set theoretic development also share the goal of providing a comprehensive foundation for mathematics. Let's now pause to reflect on what this goal does and doesn't involve.

[60] Ontology is the philosophical study of what there is; one's ontology is the sum of what one considers to exist. (Thus, 'ontology' is a subheading under 'metaphysics'.) My point is that the ontology that resulted from the Frege–Russell line of development—cumulative type theory—and the ontology that resulted from the Cantor–Zermelo line of development—the cumulative hierarchy of sets—are, in the end, practically indistinguishable.